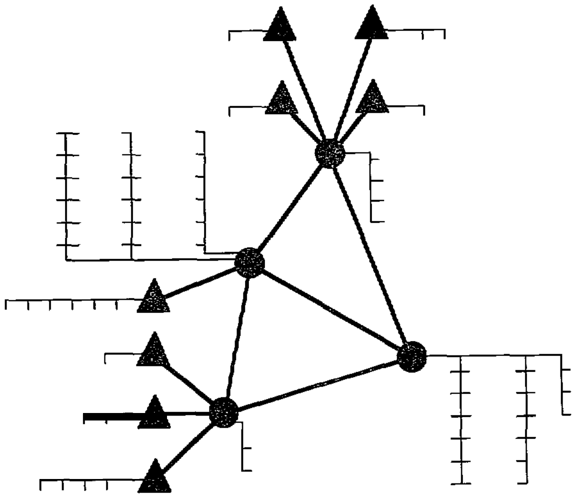


Summer 1991

The Telecommunications Review

A Publication of the Telecommunications Specialty Group

Advanced Information Systems Division

MITRE

McLean, Virginia

Table of Contents

Introduction

Henry M. James, *Editor*
Jerry Feinstein, *Associate Editor*
Victor Hsin, *Associate Editor*
Ken Wright, *Associate Editor*
Robann Spargo, *Associate Editor*

Telecommunication networks are becoming increasingly complex. They must meet the needs of sophisticated users who demand a high-level of functionality, high performance, and low cost. Telecommunications technology is changing rapidly to meet that need. The 11 articles in this issue of *The Telecommunications Review* are examples of how this technology is evolving. What follows is a brief synopsis of each article.

- Digital service networks within the federal government will be able to support high bandwidth services as new technological advancements are implemented. Frank Ferrante suggests an underlying application that could serve as a driver for requiring broadband systems on an end-to-end basis. He offers a projected architectural configuration which incorporates the new broadband technologies.

- Karen Detweiler writes a compelling case in support of Digital Audio Broadcasting (DAB) within the U.S. If DAB is more cost effective and efficient than traditional signal broadcasting, why hasn't it been implemented here? She discusses the issues and offers solutions to the problems surrounding DAB.

- Beth Foreman describes the methodology and analysis used in assisting the Patent and Trademark Office in the long-term development of the Automated Patent System. Of the three implementation alternatives, Beth recommends only one.

- Call length, frequency, and location determine telecommunications pricing. In order to find an optimal pricing algorithm. Tom Fowler uses an abstract space approach and compares the optimal result with some typical non-optimal algorithms.

- With the evolution of telecommunications networks, there is a need for interoperability between systems. Names and addresses are used in communicating with other systems. Having a centralized registration authority will eliminate the problems associated with duplicate names and addresses. Victor Hsin and Frank Ferrante discuss the administrative and technical aspects of the Government Open Systems Interconnection (OSI) Profile (GOSIP) registration implementation.

- As the federal government transitions to FTS2000, the type of data a network carries determines which services will be needed by that network. Bill Kelly uses a typical agency example to outline a framework for service selection.

- Network Management incorporates planning, maintenance, operation, network administration, and network control capabilities. As telecommunications networks become more complex and interdependent, the need for a standard interface, OSI, becomes a necessity in order to integrate multivendor equipment and multimedia services. Kris Krishnan explores the issues and technology involved.

- Users of the Intelligent Network (IN) concept would be able to design personal telecommunications services, to implement networks independent of services, and have equal access. The IN concept is described by Vladimir Nikanorov and Therese Metcalf, as well as current standards and problems with its implementation within the United States, Europe, and the federal telecommunications environment.

- Is GOSIP needed? What products are/will be available to implement GOSIP? What transitioning techniques should be used? What is the best strategy for implementing GOSIP? Richard Nieporent addresses these and other issues in his article on GOSIP transitioning.

- Isadore Schoen compares frame relay, a fast packet switching technique, with X.25 and illustrates the superiority of frame relay. Since BISDN and SMDS are not yet currently available, Izzy concludes that frame relay is the best interim solution.

- The Designer's Associate is a software package that works in conjunction with the user to produce a network design writes Ken Wright. Through a process called *propose, critique,* and *modify,* the user can then refine the proposed network design as appropriate.

We hope that you will find these articles interesting. If you wish to submit papers for future issues of *The Telecommunications Review,* please contact Henry James at (703) 883-5404.

Federal Government Broadband Networking Environment

Frank E. Ferrante'

The federal government's digital service networks, as currently configured, are capable of supporting data rates of 1 to 10 megabits per second. Based on the trends for new technology in both the local and interconnecting backbone network areas, data rates in excess of 1 gigabit per second will be supportable in the near future. This paper addresses several uses for this high bandwidth service offering, focusing on high resolution image transmissions, and forecasts a possible architectural configuration that could evolve as the federal government moves towards consolidation of voice and data networks.

Background

Technology is considered an enabling function for the future environment of federal government telecommunications systems. It is through the application of new technology that the federal government anticipates its plans for improved operational efficiency becoming a reality. Today's federal environment incor porates a number of programs which support local as well as inter-city and international telecommunications services and equipment. These include the Federal Telecommunications System 2000 (FTS2000), Washington Interagency Telecommunications System [WITS), Aggregated Systems Procurement (ASP), the Department of Defense's Military Network (MILNET), and others specifically associated with requirements unique to individual agencies. The following paragraphs describe highlights of the current narrowband federal telecommunications architecture evolution to its potential future wideband environment.

Evolution

Today's federal telecommunications networks, as illustrated in Figure 1, offer predominantly voice-oriented capabilities [1]. For voice service, the networks, like those of the public sector, are ubiquitous in geographic coverage and accessibility. Data- and packet-oriented services, on the other hand, generally have been provided either as offerings under the voice operations or as independently arranged contracts. These tend to be proprietary in form and limited in interoperability across programs. Figure 1 illustrates an example of the structure associated with current configurations. Voice and data services are shown separately except in cases where modem or modem pools are used. Local area networks (LANs), operating at 1 to 10 megabits per second (Mbps), are interconnected through either routers or gateways with host and server systems directly interfaced to carriers over dedicated circuits rather than switched directly. Interconnections to the local

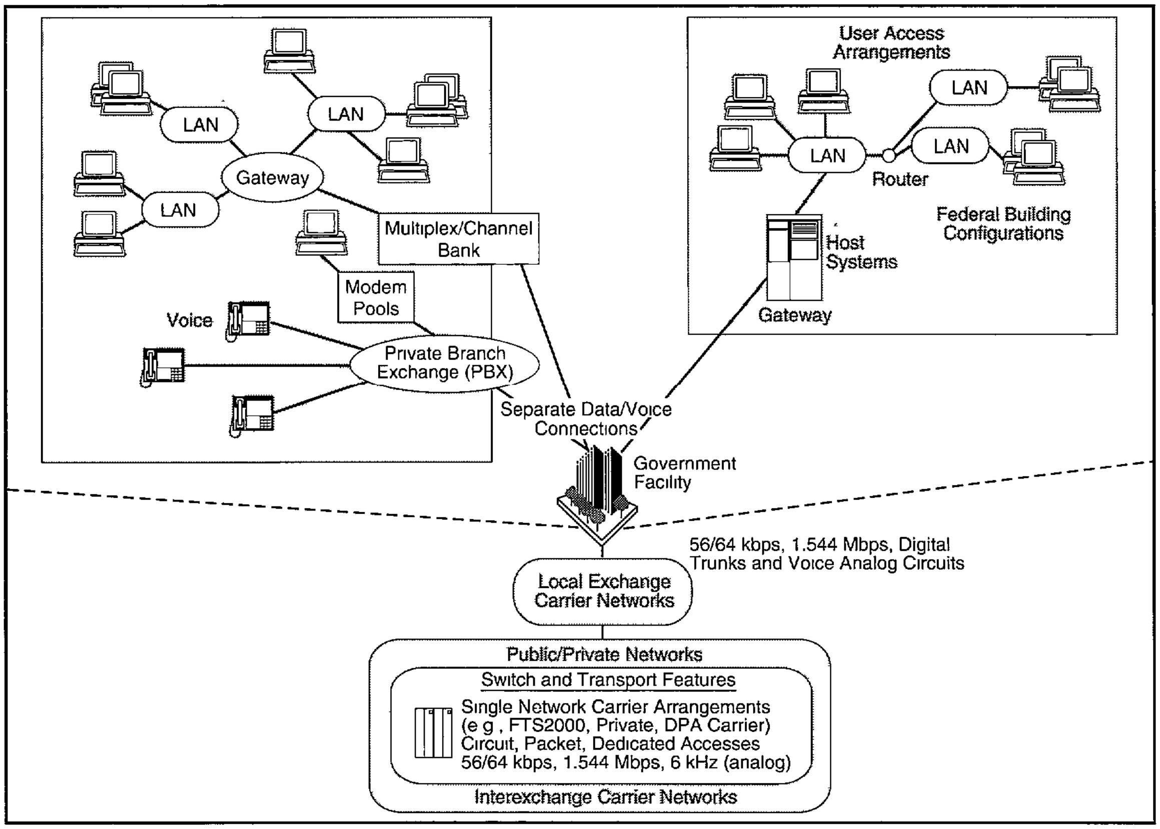


Figure 1. Current Telecommunications Network Configurations
Illustrates the relatively low speed data access and interconnectivity arrangements present in today's federal government networks and the separation of voice telephone systems from the usual digital arrangements.

carrier or directly to long-distance interexchange carriers are still a mixture of digital and analog. For most large metropolitan areas, digital service offerings are increasing; however, the market for analog will not disappear due to the prevalence of customer premises devices which depend on it for their operation.

Evolution planning for the federal government's telecommunication system has focused on reducing the incompatibilities that exist between separately procured data systems, consideration of the economics of integrating services (voice, data, and imaging), and introduction of new technological advancements and improved standards into the networks as they evolve. One direction taken to support this focus is the mandating of the Government Open Systems Interconnection Profile (GOSIP). The introduction of Intelligent Network (IN) concepts, which promote more personalized service control for users, is also being considered. Broadband services and their corresponding systems and standards have been of primary interest.

Table 1 lists transmission arrangements that represent the leading alternatives within the future telecommunications environment. The alternatives supporting broadband switched

Table 1. Transmission Alternatives

Dedicated Transmission Service
Low-Speed Data Dial-Up and X.25 Packet Services
Frame Relay
Integrated Services Digital Network (ISDN)
Broadband ISDN (BISDN) – Asynchronous Transfer Mode (ATM) - Cell Switching – Synchronous Optical Network (SONET) Framework
Switched Multimegabit Data Service (SMDS) – 802.6 Service Metropolitan Area Network (MAN) – Fiber Distributed Data Interface (FDDI) Local Area Network (100-200 Mbps) – DS-1/DS-3 Access
High Performance Parallel Interface Switching (HIPPI)

operations are of primary interest in this paper. Justification will be needed for procuring the tremendous bandwidth options being offered as a result of the introduction of Broadband Integrated Services Digital Network (BISDN), Asynchronous Transfer Mode (ATM), High Performance Parallel Interface (HIPPI), Fiber Distributed Data Interface (FDDI), and Synchronous Optical Network (SONET). HIPPI represents one of the newest offerings to be announced in supporting broadband operations [2]. Its parallel processing switching supports gigabit-per-second data transfers. To date, one of the applications described for HIPPI is the interfacing of supercomputers with disk access, file servers, and high-speed LAN gateways. ISDN makes possible the integration of voice, data, and imaging services using standard digital interfaces. BISDN extends these to higher data rates, using ATM to switch the broadband traffic at SONET-based rates. To justify the introduction of these new technologies. economic choices will have to be made based on the demand for services, such as imaging, high-speed supercomputer applications, and higher speed LAN interoperations. The following paragraphs describe the potential roles that imaging, supercomputer technology, and interacting higher speed LANs may have on future federal telecommunications architectures, given that demand for services increases as expected and that cost for systems become economically justified.

Broadband Services

Imaging includes both video as well as facsimile (Fax). Imaging is the bit-mapping of video-displayed information. Imaging involves compression and decompression, video display processing, object recognition, and mapping. Communication of Fax images, invented in the 1930s, has been exploding onto the current market. As an example of its potential, the relative growth of Fax equipment sales in the public and federal government domains outdistanced the universally popular electronic mail (E-mail) services revenues by almost five to one since 1985 and are projected to double again by the year 1999 [3]. Several factors have contributed to this growth relative to E-mail. These include: low cost of Fax equipment (relative to E-mail requirements), the introduction of a sub-minute per page Fax transmission, simpler user interface of Fax systems (familiar telephone usage), and finally, international standardization of equipment (leading to vendor interoperability).

Looking towards the future, in the medical imaging area a color or gray-scale image could use as many as 403 megabits (Mbits) of information per image (based on a 4,096 x 4,096 image size with maximum contrast of 24 bits per pixel) [4]. High resolution monitors today offer upwards of 2,000 x 2,000 image pixels. Thus, doubling of this resolution to attain the high estimate implied by the 4,096 pixels is not beyond belief in the not-too-distant future. Image compression techniques could reduce data transfer requirements by one-third. However, if a time constraint of less than a few seconds per image is imposed, this still implies megabits per second transmission rates for images.

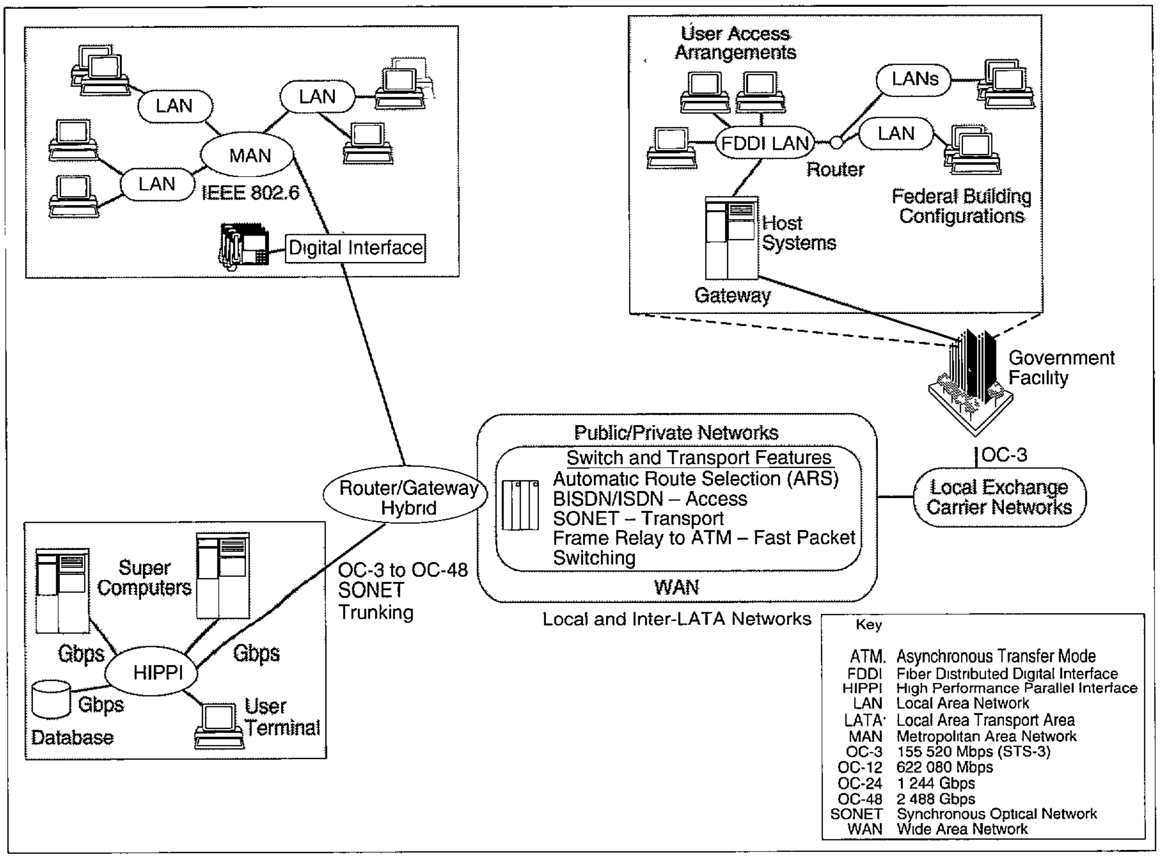


Figure 2. Federal Government Broadband Architecture Forecast

Illustrates the potential broadband interconnectivity arrangements that could be provided in the near future with an all-digital network architecture. Eliminates the separate voice and data arrangements and provides for broadband transmission services in a backbone environment which complements the need for high data transfer rates between broadband LAN systems.

Considering the rate at which Fax equipment is selling, with the addition of color, gray-scale, or clustered black and white Fax configurations, this service represents a potential leader in supporting demand for future broadband offerings. For the all-digital environment of the future, this implies that a universal switched digital service would be needed.

For federal government telecommunications systems, the remaining possible drivers for broadband services appear to be enhanced video, conferencing, supercomputer interoperability applications, and increased interconnectivity of broadband LAN operations. The relative duty cycle of these services will play an important role in confirming the demand. To date, very little data verifying total broadband requirements have been compiled.

Future Broadband Architecture

Figure 2 illustrates a projection of the federal government's broadband telecommunications architecture as envisioned over the coming decade, given that the demand for broadband services as described above becomes a reality. Illustrated are the introduction of HIPPI, FDDI, metropolitan area and wide area network (MAN, WAN) configurations, and a broadband network back-bone operations at SONET data transfer rates.

References

1. Bennington, Bernard J., 1989, *Beyond FTS2000: A Program for Change,* National Research Council, Washington, DC.

2. Hansen, C. and S. Tenbrink, 1991, "The Impact of Gigabit Network Research on Imaging," Electronics Imaging West '91, Anaheim, CA.

3. Borenstein, Nathaniel S., 1991, 'Why Do People Prefer FAX to Email?," International Federation of Information Processing (IFIP), Working Group 6.5, *Proceedings of the international Symposium on Message Handling Systems and Application Layer Communication Protocols,* Zurich, Switzerland, 3-5 October 1990, North Holland Publisher, Amsterdam, The Netherlands.

4. Kohli, Jagdish, December 1989, "Medical Imaging Applications of Emerging Broadband Networks," *IEEE Communications Magazine,* Vol. 27, No. 12.

About the Author

Frank E. Ferrante holds M.S. degrees in Electronic Engineering, and in Engineering and Public Policy. He has a dual role at MITRE where he is a principal engineer in MITRE's Advanced Information Systems Division, and *is acting group leaderfor the Telecommunications Technology group within the Telecommunications Systems department. Mr. Ferrante represents the Centerfor Civil Systems in coordinating OSI Specialty Group activities and* has *supervisory responsibility for all GSA-sponsored GOSIP registration projects. He is chairman of the Telecommunications Technology cluster group.*

Digital Audio Broadcasting (DAB) Issues and Opportunities for the 1990s

Karen Detweiler

This article is intended to inform engineering professionals about the issues and opportunities surrounding Digital Audio Broadcasting (DAB), an emerging consumer broadcasting technology. It also represents a case study of factors affecting introduction of new technology into the U.S.

What is DAB?

Digital Audio Broadcasting (DAB) is a revolutionary application of digital signal processing techniques to yield a high-quality broadcast signal with lower power requirements than conventional AM or FM broadcasts use. While the AM spectrum becomes commercially non-viable for musical broadcasts, and FM nears saturation, digital broadcasting shows great potential to deliver compact disk (CD)-quality audio over the air.

Beyond improved service, potential benefits include:

- Power and tower cost savings to broadcasters,
- Reduced potential health risks to people living and working around transmitter/antenna sites, and
- Niche applications such as broadcasting software changes to remote data collectors and processors.

Background

Digital signal transmission over the air is well-established, particularly in the military and satellite communities. It emerged because those users required highly accurate transmission of large scientific or technical data volumes over a considerable distance. Aerospace and defense applications have included telemetry and airframe pre-certification test data relays. The main difference between these forms of digital broadcasting and commercial DAB is the type of algorithm applied to minimize the data stream.

In traditional signal processing, the goal is compression-packing the largest data stream into the smallest envelope. Users need the entire data stream, since their receiving computer can process all of the data it gets. DAB is different (see figure 1). Unlike the computer, the human ear can only perceive sounds in the 20-20,000 Hz range. Within that range, some sounds will cancel each other out, a phenomenon called *acoustical masking.* DAB does not try to emulate the data stream produced by a source such as a guitar. Instead, DAB only emulates the way the human ear perceives the guitar's sound. This emulation implies using analysis and coding compression methods that differ from other digital signal processing technologies. DAB's algorithms are not derived from probabilistic mathematics alone; they also come from psychoacoustic research on how people perceive sound.

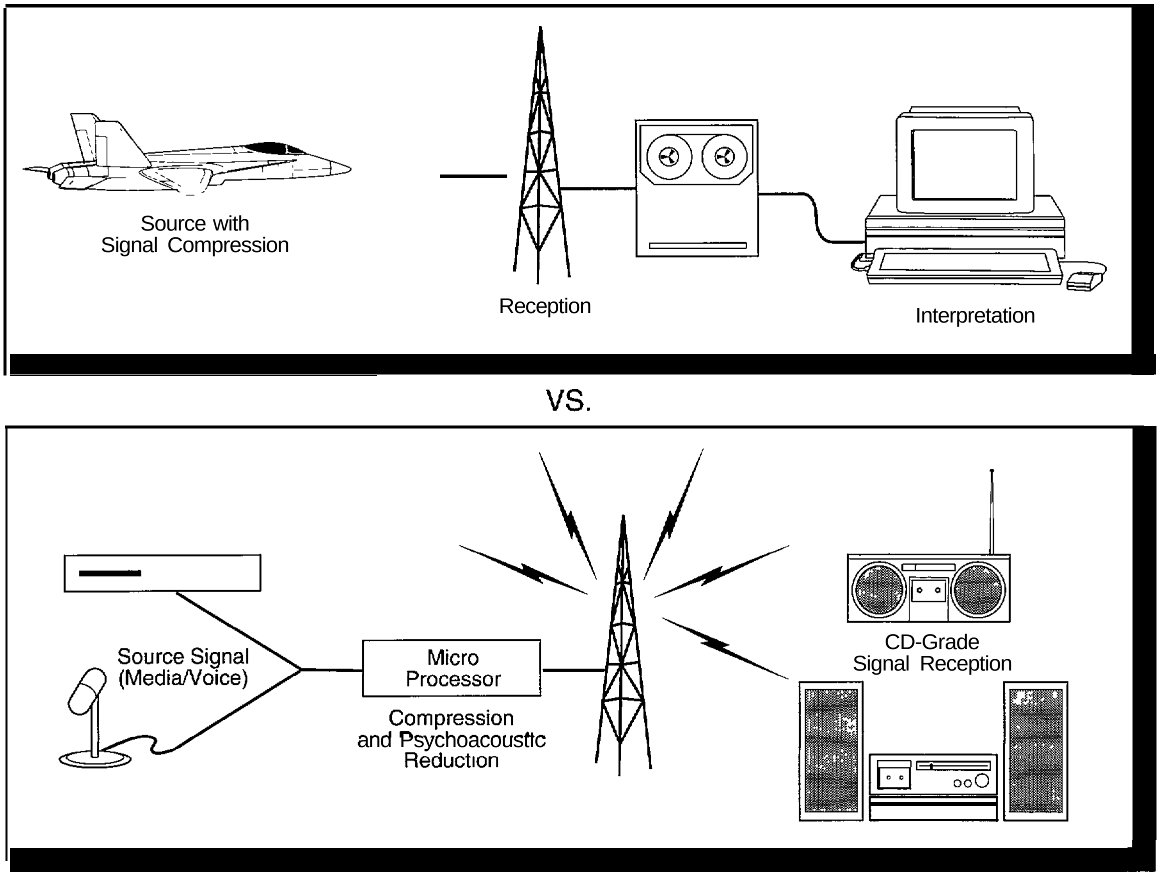


Figure 1. Traditional Signal Processing vs. DAB
Traditional signal processing and digital audio broadcasting use different methods to optimize signals.

Why is DAB Important?

Because DAB can create CD-quality sound at lower power consumption than can conventional broadcasting, DAB's technical significance is better audio service at comparable broadcaster cost. In addition, the way spectrum could be allocated for DAB could enable more efficient use of spectrum, a finite natural resource.

However, DAB also presents a case study that highlights vital issues of U.S. technical leadership in emerging technologies. The audio engineering dynamic in research and development has moved overseas, following audio component manufacturing. DAB is an opportunity for U.S. resurgence in both development and production, an opportunity for new multinational standardization, and an opportunity to increase efficiency in spectrum allocation and use.

Understanding DAB is particularly important now, as national representatives prepare positions for the World Administrative Radio Conference-1992 (WARC-92). WARC-92 will address international standards for DAB. Topics will include target spectrum allocations, coexistence with other services, types of algorithms, and international coordination. Service areas, the degree of intelligence allocated to receivers, definition of station blocks, and multinational broadcasting will be discussed. While WARC-92 cannot answer all of the questions, it will set a baseline and tone for the coming decade.

The State of DAB

Several competing approaches to DAB implementation have been proposed to the Federal Communications Commission (FCC). The FCC believes that a standard is in order, but lacks U.S. experimental evidence for decisionmaking. In preparation for WARC-92, the FCC does have an Ad Hoc Committee with subgroups addressing DAB issues. Meanwhile, other countries are moving ahead with the technology. The most visible initiative involves broadcast groups and technologists working together in a European Consortium, the Eureka 147 project.

Eureka 147 represents a $48 million, four-year investment in DAB. The Consortium is testing a prototype with the British Broadcasting Company in England, where it is easier to get spectrum allocated than in the U.S. Canadian tests are also underway; they may yield powerful stations on the border, attracting U.S. listeners away from U.S. stations.

Eureka's most distinctive feature is MUSICAM, its digital audio compression scheme. MUSICAM elegantly removes psychoacoustically masked sounds. As Steve Crowley of duTrei1, Lundin, and Rackley notes, "MUSICAM can also transmit extra information to a digital readout such as call letters, song titles, or possibly even, commercial messages." Preliminary findings suggest that these techniques enable Eureka to reduce a typical signal 80 percent before it is transmitted. In addition, Eureka's error correction/compression schemes help reduce power requirements. With Eureka technology, stations now operating at 50,000 watts could potentially provide a comparable level of service on less than 2,500 watts.

Eureka's investment matches its high sights; the Consortium seeks to create a single international DAB standard. With a single European market after 1992, and developmental buy-in from Germany, France, and Great Britain in hand, Eureka appears positioned as the European standard. As the first system to reach marketability (and before WARC-92, at that), it may create a standard for non-European nations as well. In world technology competition, Eureka constitutes a strategic European product to counter Japanese and Pacific Rim ascendancy in both innovation and production.

U.S. companies like Stanford Telecom are pursuing DAB. However, they operate at far lower levels of capitalization and staffing than the Eureka team enjoys. U.S. approaches vary (see figure 2). Some emphasize local station-owner control by copying today's terrestrial broadcast model. Others emphasize satellite-based broadcasting, which could cover the continental U.S. with a single signal in as little as three zones. The satellite-based model appeals to many people in remote, low-population density areas, because it could provide special programming like National Public Radio where markets and terrain cannot support a station today. It concerns current station owners, though, since it could increase the number of competitors in their local markets, provide an attractive environment for national advertising accounts that sustain local stations, and dilute listenership across alternative formats. A third group of DAB proponents advocates a hybrid approach that uses local towers to compensate for propagation loss (such as building and tree attenuation).

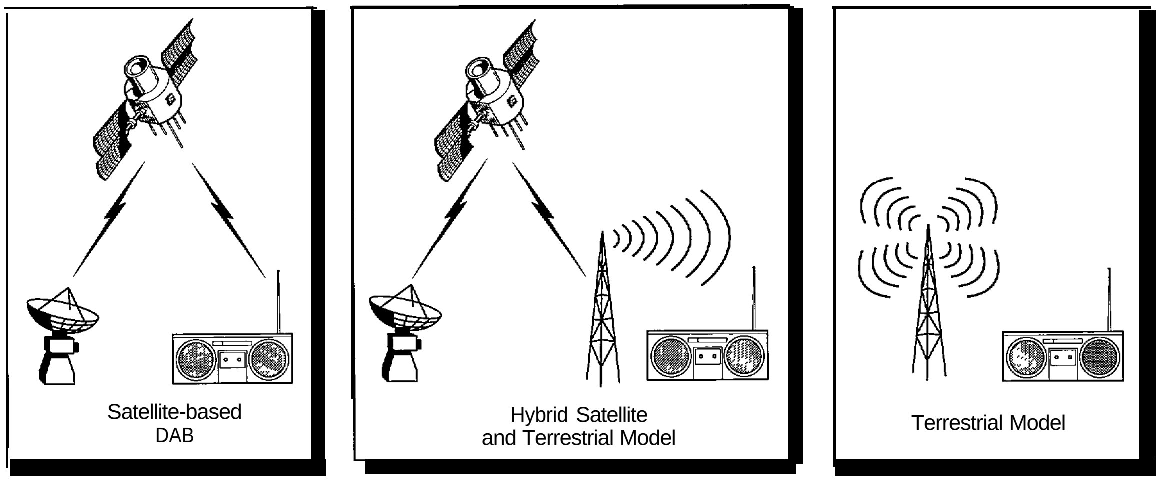


Figure 2. Options for DAB Implementation
The three prevailing options for DAB implementation are satellite-based, hybrid satellite and terrestrial, and "pure" terrestrial models.

Each of these approaches presents technical advantages, depending upon where DAB is placed on the spectrum. If spectrum is allocated to DAB in a range attractive for satellites, satellite CD radio will benefit. If spectrum comes from current TV allocations, terrestrial approaches will be more viable. The problem is a classic chicken-and-egg issue: investors are reluctant to fund a technical solution without knowing where it will reside, and they are unlikely to push for spectrum allocation unless the technical solution is demonstrably robust in that range. Meanwhile, regulators will not allocate spectrum even for testing unless there is a solid technology with strong financial backing to fill it. Gaining allocation takes time, money, and clout.

As a corollary, interest groups will support spectrum allocation drives only if they perceive them to be in their interest. Cable companies will seek over-air-to-satellite-to-cable connections, to minimize programming cost while maximizing market control and subscription revenue. AM and FM broadcasters will seek to deliver better service while retaining local control; they will gravitate toward terrestrial systems. Each of these groups will support DAB-if it is their kind of DAB, in the appropriate range, and preferably from someone else's allocation.

The FCC is considering three test allocation proposals, with varied political and technical implications. Table 1 lists the three options for test allocation proposals. Technical assessments indicate DAB performance would be optimal between 1,000 and 1,500 MHz. According to the WARC Ad Hoc Subgroup I3 discussion on a satellite-based DAB approach, "Frequencies above 1,500 MHz get progressively worse: at least 4.4 dB more satellite power is required for the same performance and coverage patterns" [7]. Subgroup B continues, "Frequencies above 1,900 are undesirable for DAB" [7]. Subgroup B favors Option 2. An FCC decision date remains unknown, and a test allocation grant does not guarantee that implementation would receive the same spectrum.

In sum, Eureka leads the technical race, while U.S. DAB research and development (R&D) remains undercapitalized, diffuse, and confronted by political and regulatory hurdles.

Table 1. Test Allocation Proposals before the FCC

OPTION	ALLOCATION	RANGE	COMMENTS
1	60 MHz	728-788	Provides an advantage to terrestrial proponents
2	32 MHz	1,435-1,530	Uses aeronautical telemetry; DAB's peak performance range is from just below 1,000 MHz and into this range
3	60 MHz	2,390-2,450	Serves the most diverse groups (e.g., fixed, mobile, amateur radio) and widespread applications (industrial, scientific, and medical)

What Factors Will Affect DAB Implementation Success?

The federal policy framework, implementers, standards-setters, and consumers will all affect U.S. DAB implementation. Many questions arise (see figure 3). For each sector, the following paragraphs give a first-cut assessment of questions, interests, perspectives, and possible actions.

The Federal Policy Framework

Could the U.S. national policy framework support or encompass digital audio initiatives (figure 4)? The Bush Administration's *National Technology Policy* (hereinafter called the *Policy)* document of 26 September 1990 provides the key.

The *Policy* notes that "We are in an era marked by increased international economic interdependency and increasingly stronger technological capabilities in other industrial nations." DAB bears witness to this shift. It continues, "as new products mature, the advantage shifts from the innovator to the efficient producer." Lacking the innovative edge in DAB, the U.S. could play innovation catch-up, or focus on becoming an efficient producer.

The *Policy* stresses market-driven resource allocation, and seeks to capitalize on "the strengths of our economic system more effectively to help U.S. firms remain competitive." The DAB case suggests a need to regain competitive status, not just remain competitive.

The *Policy* emphasizes private sector leadership. It asserts the private sector's responsibility to "identify and aggressively pursue potential commercial applications for technologies developed by its own laboratories, as well as by universities, federal laboratories, and foreign sources" [2]. DAB provides an immediate example. It continues, "we need to refrain from actions that might distort our basic system of free enterprise." That statement tends to mitigate against preferential treatment for technical pioneers or established broadcast licensees who might push new technology, and perpetuates today's gridlock. As the federal R&D role, the *Policy* sets up "streamline federal decision-making structures and mechanisms to eliminate unnecessary and cumbersome regulations and practices that inhibit industrial competitiveness." It thus keeps technical and financial risk squarely in the private sector, and offers only a government goal of regulatory streamlining.

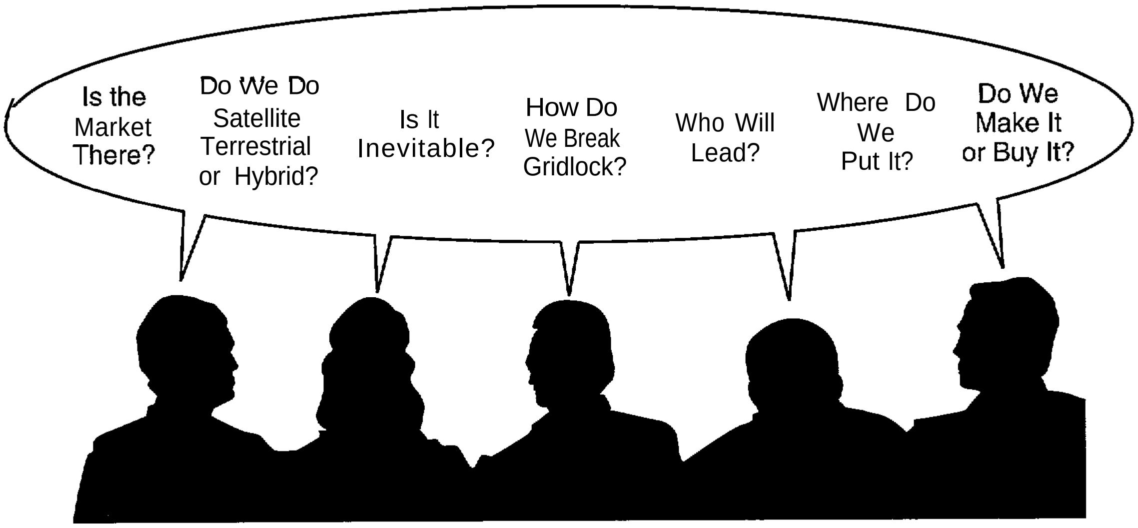


Figure 3. The Big Questions of U.S. DAB
There are many practical questions about DABS success potential.

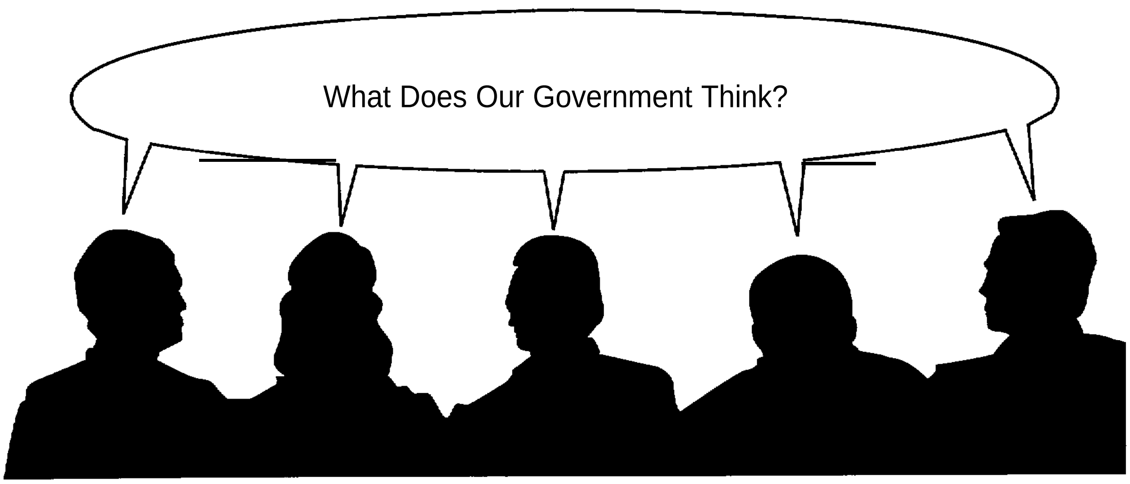


Figure 4. What Is the Government's Position?
Changing the status quo requires that we look at how flexible the current policy is.

Regarding standards, the U.S. position is to "encourage increased U.S. participation in multilateral international standardization efforts through the standards activities of the National Institute of Standards and Technology" (NIST) [9]. This position connotes special interest in gatherings, such as WARC-92, though FCC-more than NIST-is looking at DAB. Even so, products (not committees) increasingly set international standards and provide competitive advantage. Committee work lags, while products create de facto standards with revenue streams.

Clearly, the high-level policy framework to enable U.S. DAB service exists. International cooperation, technology leveraging, and competitive innovation and production are prominent themes relevant to DAB. Whether the *Policy* cohesively migrates into the regulatory system remains an open issue. Further, DAB presents a case of innovation and manufacturing advantages abroad, with U.S. firms seeing little incentive and considerable risk in innovation. In this context, a production stake may be the best position U.S. free enterprise can achieve under current policy.

The Implementers and Standards-Setters

DAB's prospects and impact depend on complex interactions among many implementers and standards-setters (figure 5). Let's look at the interests and orientations of some players who will affect DAB implementation.

Professional Associations: IEEE broadcasting groups are technically minded, and leave the lobbying to trade groups. However, having been burned by the AM stereo fiasco, which emerged without a clear standard for either transmitters or receivers, even the IEEE may be conservative in addressing a new technology framework for the 1990s. While technical interest exists, endorsement of any approach is unlikely.

Trade/Industry Associations: National Association of Broadcasters (NAB) groups are principally owner/investor-driven, and represent about half of the stations in the U.S. On 29 January 1991, the NAB Board unanimously approved the Eureka 147 technology. This endorsement shifts the NAB from impartial assessor to proponent status. In moving toward licensing Eureka's system, the NAB may be attempting to block satellite, broadcast network, and cable companies from applying the technology, and keep U.S. implementations terrestrial. However, endorsement is just one small step in delivering DAB service to consumers, and the NAB does not represent all broadcasters. If DAB appears as a threat to an established broadcaster's economic base, he will fight it just as the recording industry lobbied to block digital audio tape (DAT). Only if the government regulatory authorities could grant preference to established broadcasters would they be likely to embrace new technology enthusiastically and incur the concomitant cost risk.

The nature of U.S. equal access laws makes overt preference toward existing licensees unlikely; instead, the timeliness with which DAB license applications are submitted and processed would probably become the margin of economic advantage. Since professional broadcasters may be better informed (and better capitalized) than external interested observers, current licensees can probably dominate the first wave of applications. Of course, being first or best is not as important as being observant and consistent; later licensees may have a higher business success rate as they learn from DAB pioneers.

Among those committed to the DAB concept, investors and potential licensees may perceive Eureka as a lower risk, faster-to-market strategy than waiting for a U.S. version of DAB.

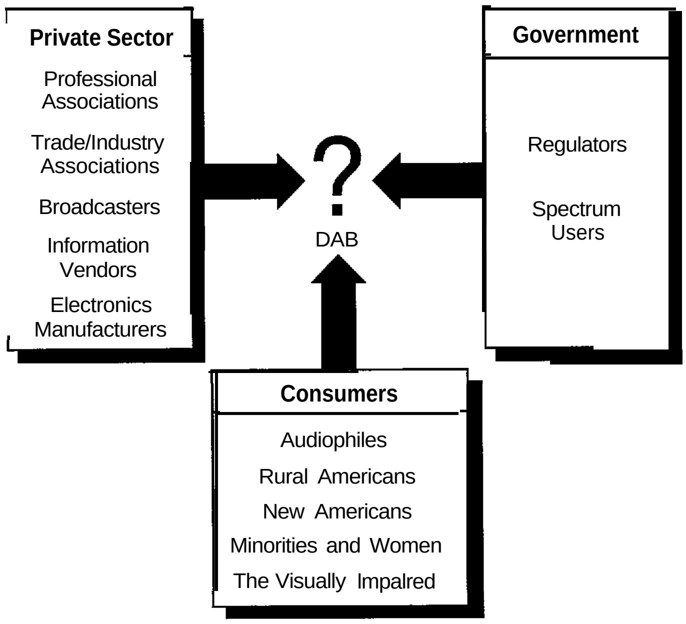


Figure 5. Who Will Affect DAB Implementation?
Three sectors. each comprising many constituencies, will ultimately shape DAB's success in the U.S.

Broadcasters: We touched on local broadcasters' interest in local control and terrestrial service above, as well as networks' interest in serving larger markets directly via satellite. Other broadcasting players merit consideration, too. Voice of America and other internationally oriented broadcast organizations have considerable expertise in today's technologies: AM, FM, and shortwave. U.S. groups have an interest in projecting their U.S. programming abroad, just as other countries want to continue broadcasting into the U.S. If other parts of the world in which we have strong economic and political interest (such as Europe and the Pacific Rim) migrate to higher quality DAB systems, international broadcasters will have little choice but to follow. They may not have the luxury of advocating an implementation, and would adopt the standard of the nations they serve. Clearly, they would benefit from compatibility if the U.S. adopts a multinational standard or the world achieves a single standard.

In terms of cost to the broadcaster, the difference between a new FM and new DAB station is likely to be small. Since staff costs and load are comparable, there may be little economic difference in operating costs. The cost model will differ, though; a Eureka-style implementation would not allocate channels individually. Instead, 12- 16 stations in a 4 MHz block of spectrum would share a transmitter and have identical coverage. Changing the ratio of stations to transmitters and evening out coverage areas would affect both the per-station capital investment required and the market definition. If current licensees were to receive preference in DAB licensing, this democratization could aid small or poor-quality station licensees, and erode the advantage now enjoyed by powerful stations. Projected revenue then drives the cost model.

Information Vendors: The broadcast industry brokers airtime. It delivers news, promotions, special programming, and music that is licensed.

Essentially, news is a loss-leader community service, like public service announcements and the Emergency Broadcast System. Stations pay profit dollars to networks for tie-ins to news services, and only rarely carry the expense of local news staff. Promotions, programming, and music are ways to attract a public to hear sponsors' paid commercials. From this perspective, these behind-the-scenes information vendors have special interest in DAB since they exist to sell compact disks, video tapes, and other formats of the same information from the same creative talent. These information vendors are increasingly under foreign control or ownership, which may buttress their interests in international DAB standards.

Without locally controlled radio broadcasting, vendors like local advertisers would have no place to air low-cost audio spots; they would be pushed into declining market print ads or higher production cost television ads. Without station owners' autonomy in choosing formats, program and music vendors would lack outlets to stimulate sales demand for their products. Information vendors will likely seek a broadcasting mix to maximize potential listenership through mass-market and innovative niche outlets, using old and new technology to reach different consumers.

Regulators: Since the broadcast spectrum is a bounded, saturated, and highly politicized resource, the FCC has little interest in stirring already muddy regulatory waters to reallocate spectrum so DAB can be implemented. However, both the FCC and the National Telecommunications Information Administration (NTIA) recognize a need to accommodate new technology. They realize the significance of WARC-92 on the international scene, and seek to retain their domestic clout in its wake. Creation of the Ad Hoc Subgroups to prepare for WARC confirms this interest. Concurrently, at least one FCC Commissioner has encouraged DAB's principal U.S. venture capitalist, Ron Strother of Strother Communications, Inc., to continue nudging DAB forward.

Will regulators be pro-active or reactive in addressing the new technology? Reactive regulation characterizes U.S. regulatory history, but could kill U.S. initiatives and result in late adoption of a foreign-originated DAB standard.

Existing Spectrum Users: The intelligence, military, and drug enforcement communities have considerable interest in the forthcoming spectrum allocation debate. The Aerospace and Flight Test Radio Coordinating Council (AFTRCC) and the Interdepartment Radio Advisory Committee (IRAC) represent other regulatory interests, whose positions may be summarized as "DAB is fine, just not in my allocation." Spectrum is valuable; once lent, it may never be returned, so these communities are loath to part with any spectrum, even if underutilized today. To accommodate their interests and free spectrum for DAB allocation, several options are apparent:

- Compensate them for relocating their users to other portions of the spectrum already allocated to them, or to new allocations, or

- Provide a rental or royalty-fee arrangement to enable those organizations to be compensated over time, and create a "Hong Kong" plan granting new DAB licensees a license to use certain allocations for a lengthy period.

Electronics Manufacturers: U.S. defense contractors, retooling and adapting digital signal processing technology, could grow DAB implementations, including receiver manufacturing. They possess adaptable technical talent and custom chip development capability. In manufacturing, as in broadcast licensing, the strategic advantage is with the "haves."

Internationally, the Eureka consortium members have the manufacturing lead, with working prototypes and planned refinements. If they license algorithms and/or chip designs, the Japanese and Koreans could move into quantity production quickly. Yet, such licensing is unlikely in the near term; the Europeans see Eureka as a strategic advantage over Asian technology, and probably intend to resuscitate European manufacturing with closely held designs. U.S. manufacturing interests could seek licenses to manufacture Eureka products, but may have to support a U.S. Eureka standard implementation to gain licenses. This strategy could re-establish the U.S. manufacturing position and provide employment. It is also consistent with U.S. *Technology Policy* to promote "harmonization of regulations and standards for products and processes with our major trading partners" [2].

Consumers, or, Is There a Market?

Will consumer demand encourage DAB investors? Today's general consumer market reflects rapid shifts and great unknowns. What will be the consumer's priorities? In a cost-conscious market, will current technology suffice? Will discretionary dollars (1) exist and (2) be directed toward quality and value criteria? Will cocooning yuppies, the graying of America, fitness fanaticism, and increasingly long commutes create a strong pull toward the new technology? Or will the economic slowdown and aging baby boomer conservatism decrease the acceptability of new technology? If the world standard originates abroad, will the U.S. experience the "Not Invented Here Syndrome" with isolationist technical, economic, and cultural effects? More importantly, will these factors in today's market still be relevant in five years, when DAB could be widely fielded?

How format flexible are consumers? Despite the higher unit costs of compact disk audio, 1980s consumers demonstrated a willingness to move rapidly out of less expensive, less durable long play (LP) albums. According to the DAB-Eureka Consortium's FCC filing citing statistics of the Recording Industry Association of America, "from January to June 1990, 132 million CDs were sold in the U.S. versus 5.6 million LPs- a ratio of nearly twenty-four to one." A key issue in the 1990s is whether this flexible attitude toward format will persist, and carry over into related technologies like DAB. The degree to which all other interest groups address DAB technologies hinges upon whether the financial community believes that a potential market exists, how large it might be, and how rapidly it might establish itself in the U.S.

Surveys provide some preliminary insights. An A.C. Nielsen market survey reported by John Abel to the NAB Group Executive Forum suggests that consumers will remain format-flexible, if the perceived increased quality of service outweighs the user costs. In assessing customer willingness to pay for CD-quality sound transmitted through domestic cable, the Nielsen survey indicated that 14 percent of consumer households were "Very Likely" to subscribe to even a pay CD-quality radio at $7.50 per month" [8]. Abel's presentation included findings that 88 percent of Digital Cable Radio's customers feel that the new CD-quality radio-over-cable service met or exceeded their expectations, and more than two-thirds say it is their most-used stereo component. These surveys trigger further questions about DAB market robustness and positioning.

- *"Which consumers would gain from DAB?"* is an important, related question. Consumer electronics have long reflected a trickle-down effect, starting with audiophiles and gadget junkies in the higher socioeconomic brackets. DAB technology can be expected to supplant FM among high end users, as FM displaced AM. How rapidly is the open question.

New Americans: Today, AM stations are failing regularly, so community licenses are relatively inexpensive. The receivers are universal; their costs trivial. The only growth in AM markets appears to be in tightly clustered recent immigrant communities. Depending on the pervasiveness of English in near-term U.S. economics, and the willingness of these communities to assimilate, the best case economic lifespan of these stations may be one generation, mirroring the lifespan of countless new-arrival foreign language newspapers in prior generations. Could these new Americans afford to buy, run, and grow FM or DAB outlets? Their immediate prospects are poor.

Minorities and Women: To quote FCC Commissioner Barrett in *Communications Lawyer,* "It has long been recognized that diversity of media ownership is favored by the First Amendment and is one of the main components embodied in the FCC's public interest mandate" [1]. To that end, U.S. minority group representatives have already filed comments on DAB with the FCC. Though they have received little media coverage, their point remains: only l-2 percent of all U.S. broadcasters licenses are held by minorities. The filers maintain that the reallocation of spectrum and the DAB station licensing process present opportunities for greater minority ownership. With current conditions, the only practical way to get a license is to buy an existing station, and few viable properties turn over. With DAB, some minority representatives see an opportunity to build a property from initial licensing, which carries a greater potential gain than buying a property and financing upgrades. An underlying current of the minority argument is that diversified ownership increases the chances that diverse formats will exist and serve more communities more effectively. A similar argument applies to female ownership. "While the FCC has stated goals to increase female and minority ownership of telecommunications properties, policies for increasing minority and female ownership have been challenged and are now pending review before the United States Supreme Court" [1]. Regardless of the rulings, ownership shares will remain a sensitive issue for the foreseeable future. Consideration of preferential DAB licensing to existing licensees may merit tempering in light of the minority and female ownership arguments.

Rural Americans: Rural U.S. consumer regions are underserved; they include mountainous and/or low population density areas. Some DAB implementations are positioning themselves for public radio service to these areas, usually involving satellite or hybrid systems.

The Visually Impaired: The consumer sectors which first benefit from new technology are rarely those that most need it. Visually impaired people, for whom braille-coded cassettes have been a great boon, would surely appreciate DAB more fully than most other citizens. While many have benefited from current audio technologies, few have migrated to CDs due to (1) expense, and (2) difficulty telling CDs apart by feel. Coupling DAB with DAT would certainly be a major benefit to the visually impaired-if the price is right.

- *How would the consumer use DAB?* To replace AM/FM? To supplement PM? To enjoy (then record DAT) broadcast performances? (This last premise terrifies broadcasters, performing and recording artists, and audio production/sales people alike.) Instead of TV? Probably not.

- *When would the consumer use DAB?* Signal quality to moving receivers is likely to be a major DAB strength over conventional AM and FM broadcasts. This factor, in turn, suggests that vehicle installations could come first, with home system units and walkman-like lightweight personal receivers following in time.

- *Would the customer turn on DAB?* Americans' penchant for the new suggests yes, providing the costs were not highly skewed above conventional audio broadcasting. The gain in broadcast quality must exceed the marginal cost of the receiver, and program format variety must exist in major markets for the technology to become widely accepted.

- *How would listeners learn about a new radio technology?* Crossover ads from local TV and print are a possibility, but without simulcast licensing, it is unlikely the radio listeners would rapidly adopt the new options. The assertion that new service justifies new consumer costs must be supported with the formats the customer wants, delivered at a demonstrably better level of quality.

How Might DAB Come to the U.S.?

A coordinated strategy to bring DAB to the U.S. is increasingly unlikely as regulatory

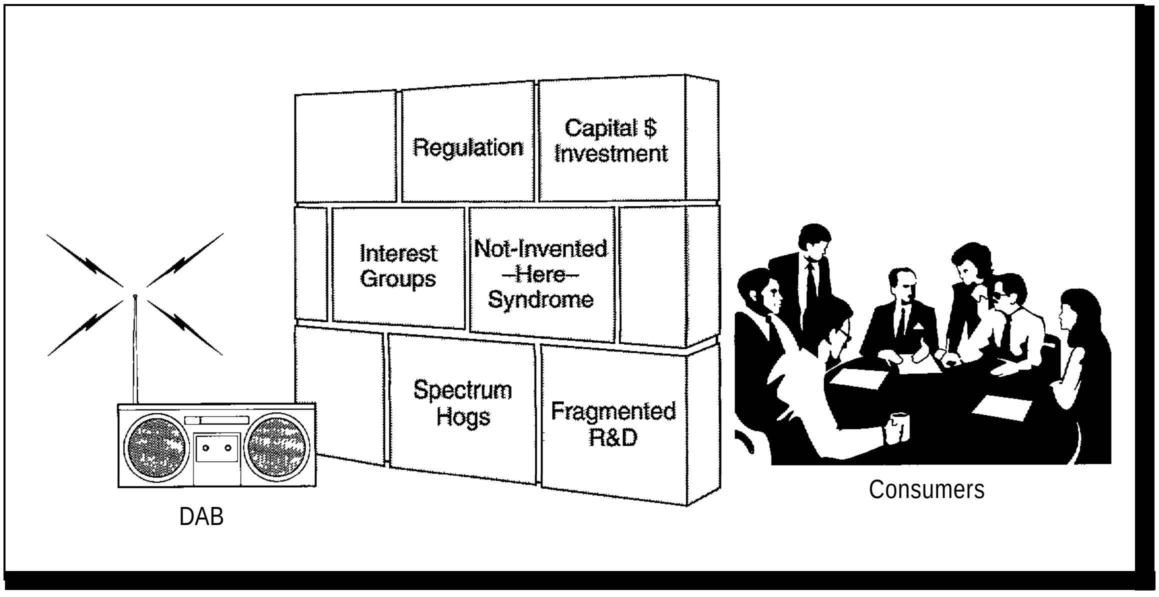


Figure 6. Obstacles to DAB Reaching U.S. Consumers
The U.S. status quo presents a brick wall that may keep DAB from reaching consumers

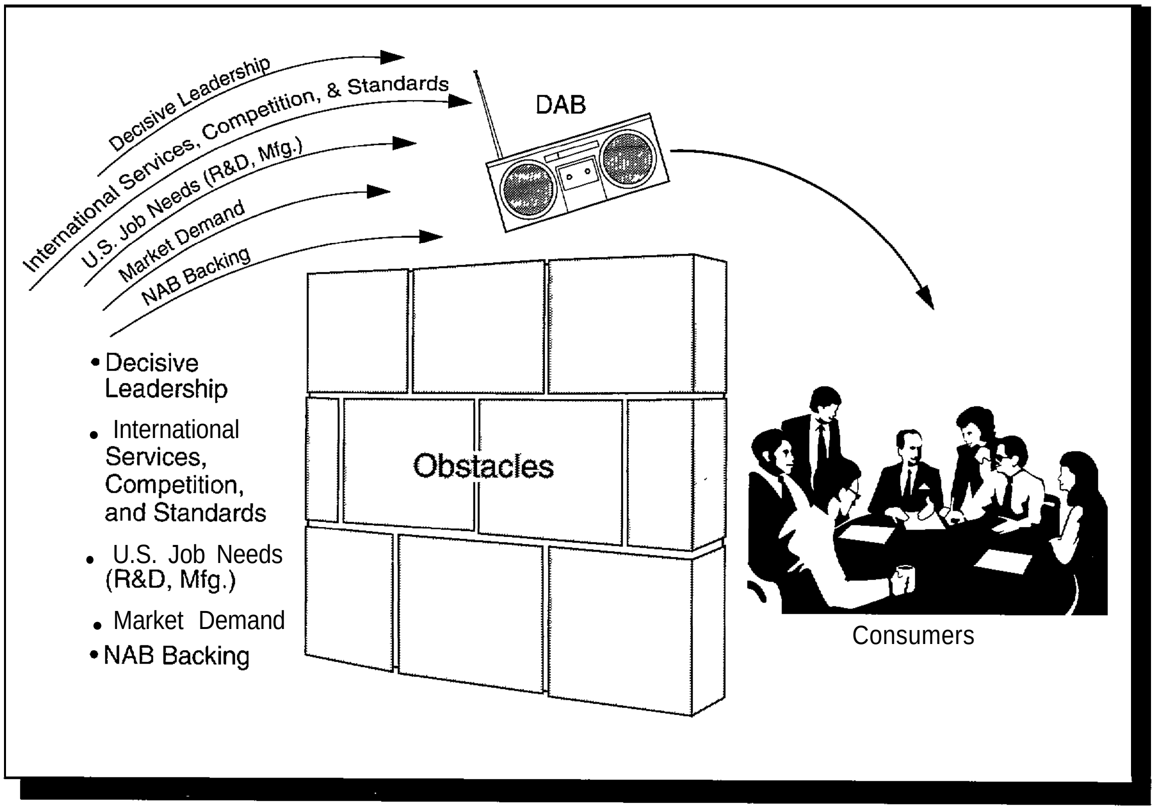


Figure 7. Forces That May Propel DAB into U.S. Consumers' Hands
Some forces may act together to bring DAB over today's hurdles to consumers.

inertia impedes granting of experimental authority to test DAB within the U.S. Numerous obstacles may block U.S. DAB service (see figure 6). A combination of internal and external forces, beginning with decisive leadership, could push DAB over the hurdles (figure 7). An incremental strategy by well-financed proponents remains possible. Because the NAB endorsed a foreign-originated DAB standard, it may have signaled the U.S. manufacturing community a clear intent to proceed, possibly rejuvenating an "off-shored" consumer electronics industry (NB: It may have simultaneously surrendered the last vestige of U.S. leadership in audio technology). Though the NAB has taken a position, the market decision still may be a plurality of standards that recreate the "BETA versus VHS" or "AM stereo versus AM/FM" shakeouts, with costly, time-consuming, and technically inefficient results. Absent a deliberate process, DAB may come into the U.S. as people in border areas purchase receivers to hear Canadian or Mexican DAB programming.

The Next Steps

DAB technology may have the "right stuff' to define a new market, improve consumer audio broadcast service, and thereby improve the American quality of life. However, strong forces discourage introduction of new technologies into U.S. markets: changing customer values and preferences, established business interests, and complex, heterogeneous regulatory structures.

Whether the impetus is external or internal, rapid or slow, change toward a DAB service is likely. The degree of success, in terms of U.S. ownership and leadership, will depend on a confluence of changing attitudes, beginning with the regulatory community. If the current policy is extended beyond the high-leverage technologies explicitly identified in *U.S. Technology Policy* [2], DAB should receive some attention. However, mere tacit support is unlikely to enable new technology to clear regulatory, political, and technical implementation hurdles.

The U.S. has precedent for unique technical standards. Can it afford to lock in on a unique national DAB standard in an increasingly international telecommunications market? Not likely. While the post-war boom allowed the super-dominant U.S. to settle on the NTSC television display standard while others moved ahead, today the U.S. has neither the short-term technology lead nor the financial pre-eminence of that era. Likewise, the national business interest is unlikely to be well-served if DAB turns into a "BETA versus VHS" standoff. Given foreign ownership of major media production and distribution channels, a new technology could falter quickly if information vendors rejected a U.S. standard as incompatible with their national systems or the larger world market.

To implement DAB successfully, the U.S. needs a systematic, multi-sector approach to introducing new technology. Such an approach would integrate the issues and interests of many affected parties including manufacturers, broadcasters, regulators, and consumers. It would address the role of standards, national interest, technological leadership and international compatibility. Key factors will remain (1) who sets the standards, and (2) when are they set?

The strategy must include market penetration for receivers, beginning with concepts like DAB receivers optional in car sound systems, or compatible with home audio systems. It must also address the technical approaches, and accommodate varied approaches (satellite, terrestrial, or hybrid systems).

For broadcasters and regulators, the process must include economic viability factors, such as allowing AM licensees, FM stations, and local TV stations to simulcast in DAB, serving both high and low ends of the market. Cost modeling for the migration would need to show advantages for the DAB service, and DAB products. It must also be sensitive to the U.S. economy; if deep unemployment among ex-defense high-tech professionals persists, government support to transmitter and receiver manufacturers could be a mitigating option.

Additional research, demonstrations, and evaluation of competing systems are vital prerequisites to competitive U.S. standards-setting and rational spectrum reallocation. The U.S. has a small window of opportunity to shape this new technology and affect standards. Unless the gridlock among special interests, regulators, and investors can be broken, U.S. engineers will lose the opportunity to compete. While Eureka leads in the technology and maturity curves, it might be surpassed if the U.S. could act quickly and decisively. Persistent gridlock may prevent the resurgence of U.S. competitiveness in broadcast audio technology. The penalty for lost technological leadership exceeds mere prestige; it leaves U.S. station, transmitter, and receiver owners paying royalties in perpetuity to engineers abroad. Can the U.S. afford to pay audio engineering royalties that will fund enhancements abroad? National interest suggests that regaining technical competitiveness would create a cumulative investment in the future of U.S. engineering. It mandates strong, decisive, immediate action to break today's gridlock.

References

1. Barrett, Andrew C., Spring 1990, "Ownership in the Communications Industry: Opportunities for Minorities and Women in the 1990s," *Communications Lawyer*, Volume 8, Number 2, The Forum on Communications Law, American Bar Association, pp. 1, 17.

2. Bromley, D. Allan, 26 September 1990, U.S. *Technology Policy,* Executive Office of the President, Office of Science and Technology Policy, pp. 3-4.

3. Crowley, Steve, 3 June 1990, "Bringing Radio up to Date," *Washington Post*, pp. D3.

4. Crowley, Steve, 15 February 1991, "Special Report: How the Eureka 147 Digital Radio System Works," Special Report in *Radio Business Report.*

5. *Digicasting: The Sound of the Future* (Pioneer Rule Anthology), Strother Communications, Inc., Digital Audio Broadcasting, Ltd.

6. *Digital Audio Broadcasting: Comments of the DAB-EUREKA 147 Project Office,* 13 November 1990, Filing before the FCC, GEN Docket No. 90-357, pp. 7-8.

7. Reinhart, E., 8 November 1990, *Comments of Ad Hoc Subgroup B on the FCC's Second Notice of Inquiry on WARC-92 Issues,* (GEN Docket No. 89-554), pp. 14, 32.

8. Rothblatt, Martin A., and Peter J. Dolan, 13 November 1990, *Comments of Satellite CD Radio, Inc.*, Filing before the FCC, GEN Docket No. 90-357.

About the Author

Karen Detweiler is a member of the technical staff in MITRE's Center for Civil Systems Division, Information Technology Department. She holds a bachelor's degree from the University of Virginia, and an M.S. in Systems Management (Information Systems) from the University of Southern California. Before joining MITRE in September 1990, Karen worked extensively in studies and analysis in BDM international's Special Projects Division, and in software systems development at McDonnell Douglas. Her projects have included teqlecommunications systems and information systems for C^3I, medical, and corporate users.

Editors note: During the publication process, the FCC granted a "'pioneer's preference" to parties requesting spectrum allocation rule changes [GEN Docket No. 90-217]. The FCC's action paves the way for innovation of new communication techniques with fewer restrictions.

An Open Systems Interconnection Conversion Strategy for the Automated Patent System

Elizabeth A. Foreman

This paper describes the high-level analysis that the author performed to determine an Open Systems Interconnection conversion strategy for the U.S. Patent and Trademark Office's Automated Patent System. A more detailed description of the analysis can be found in MITRE Technical Report MTR-89WOO207 (see references on page 33).

Introduction

Since the mid-1980s, the United States Patent and Trademark Office (PTO) has been developing an Automated Patent System (APS) which will assist patent examiners-as well as science, industry, and the general public- in obtaining current information about U.S. and other nations' uses of technology. Such users will employ the distributed processing capabilities of the APS to search, retrieve, and display U.S. and foreign patent information in both text and digitized-image forms over the APS's high-speed data communications network.

Currently, the proprietary Xerox Network Systems (XNS) communications protocols developed for ethernet local area networks (LAN)- along with customized software developed for APS-specific implementations of the XNS protocols-support such activities; however, the Government Open Systems Interconnection (OSI) Profile (GOSIP) Federal Information Processing Standard (FIPS) and the APS's future connectivity requirements will necessitate that the APS ultimately convert to the International Organization for Standardization's (ISO) OSI protocols.

Although the APS currently supports only 50 to 60 users in a limited operational environment, the PTO plans to deploy the APS, in phases, to more than 1,500 users between 1991 and 1997. This will involve acquiring and installing additional hardware and software, adding new applications and operational capabilities, and replacing the APS's private branch exchange (PBX)-based LAN (see figure 1). Since the needs of government, science, industry, and the general public for patent information will not allow the PTO to delay, interrupt, or divert resources from these plans, any protocol conversion must be scheduled either before or after such deployments- that is, in 1991 or in 1997.

Methodology and Analysis

This high-level analysis examined three alternative protocol conversion paths which the PTO could follow to meet its requirements:

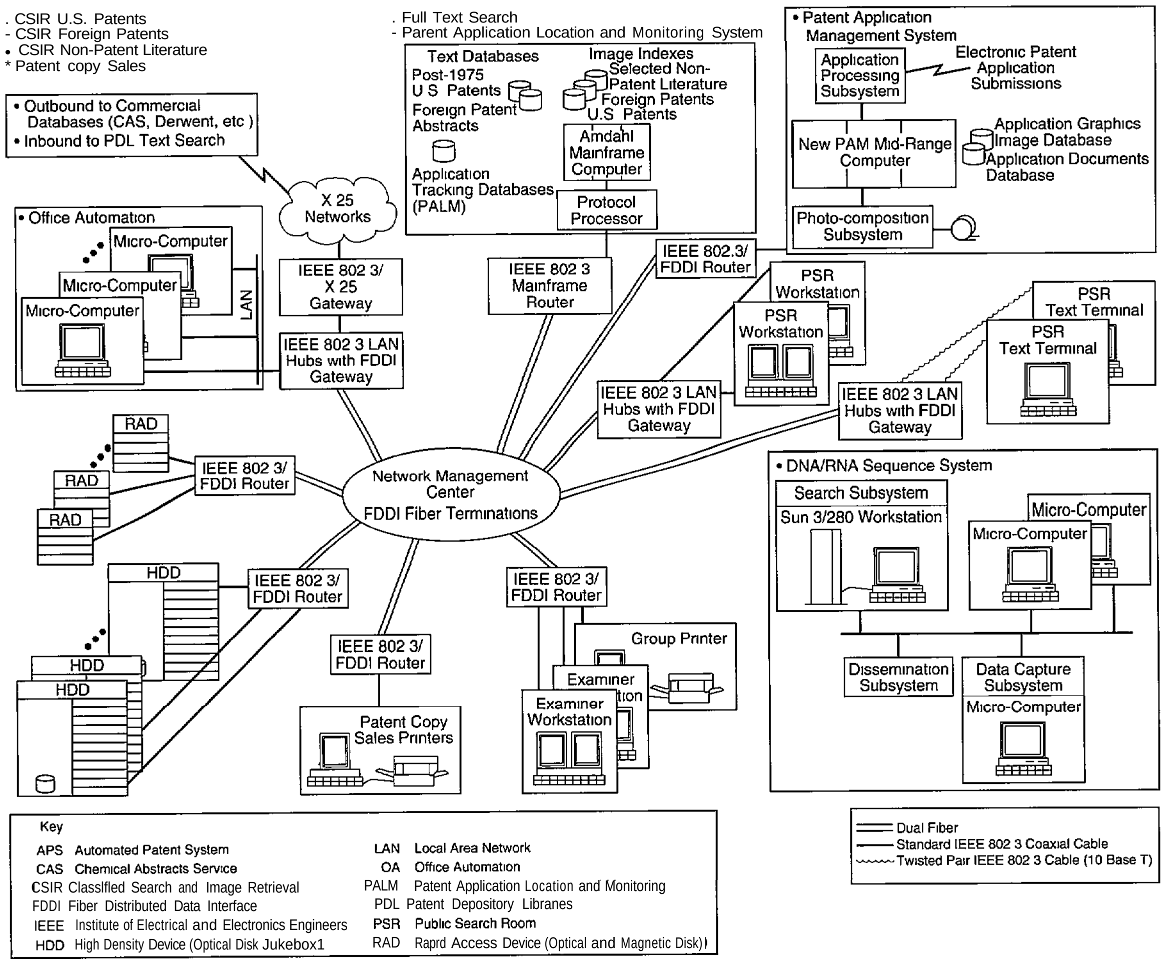


Figure 1. Overview of 1996 APS Architecture

By 1996, the Automated Patent System will have over 1,500 users as well as a variety of devices, applications, and connection characteristics which its data communications protocols will need to support.

- **XNS Alternative:** Continue to use, purchase, and deploy XNS-based equipment and protocol software until 1997, then convert to the OSI protocols and replace all XNS-based hardware and software (purchased and in place since 1986) with OSI-based products

- **DDN Alternative:** Convert to the Defense Data Network (DDN) protocols, e.g., Transmission Control Protocol/ Internet Protocol (TCP/IP), in 1991 as an intermediate protocol suite and perform a DDN-to-OS1 conversion in 1997

- **OSI Alternative:** Convert from the XNS protocol suite to the OSI protocols in 1991

Each alternative can support the data communications functions that the APS requires and meet the performance requirement for an average

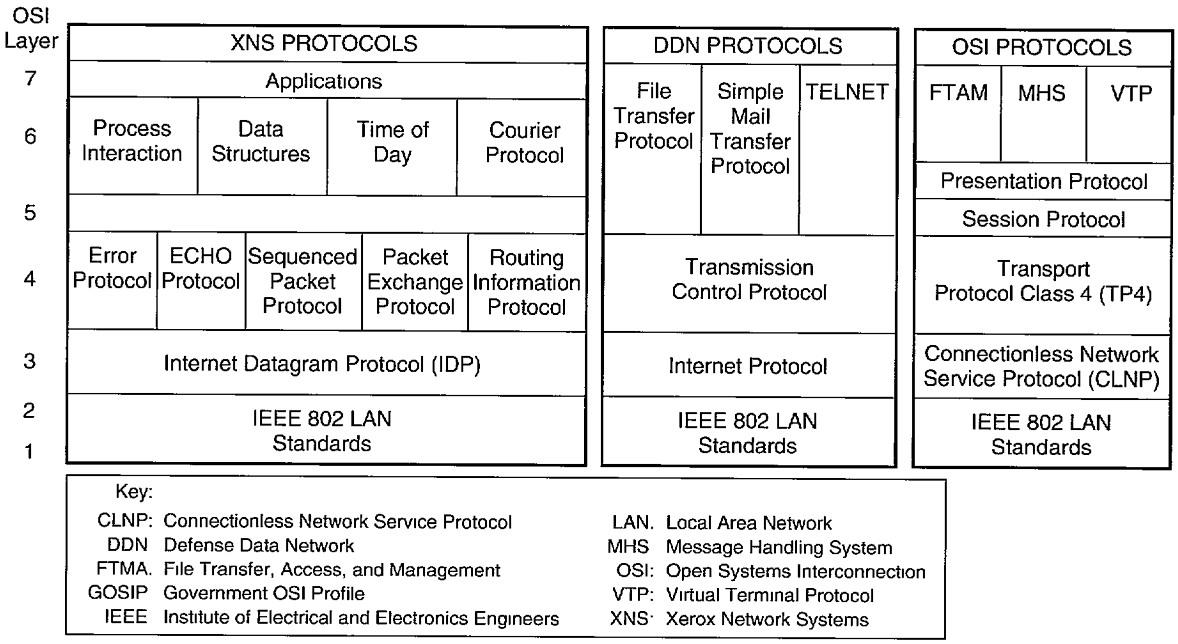


Figure 2. XNS, DDN, and OSI Protocols
The fact that there is no one-to-one correspondence of functions and services among the XNS, DDN, and OSI protocols will add to the complexity and costs of the Automated Patent System's protocol conversion.

APS single-image-transfer rate of 1 megabit per second (Mbps). Figure 2 shows the major protocols involved in each alternative as they relate to the OSI seven-layer reference model.

Four criteria with which to evaluate the alternatives were selected and ranked. The criteria, summarized in table 1, and the reasons for their ranking are as follows:

- **Maturity of the Protocol Standard:** The extent to which a protocol standard has been implemented, tested, and refined in academic, laboratory, and actual work settings:

  The PTO needs mature protocol standards to assure stable protocol operations among the various APS devices and to minimize the interpretation and integration problems, delays, and costs that are characteristic of products based on standards not yet mature. The more a protocol standard is exercised, the more confident the PTO can be regarding the stability and utility of the standard, the standard's future applicability to APS requirements, and the standard's endurance throughout the APS system life.

- **Market Conditions:** The availability, demand, and variety of products that implement a protocol standard or set of standards:

  The PTO needs commercial products that have solved problems similar to the APS's, are available from several sources, and belong to product lines that vendors are committed to support.

Table 1. Conversion Alternative Selection Criteria and Factors

CRITERION	FACTORS	EXPLANATION
1. Maturity of the Protocol Standard	Status of the Protocol Standard	e.g., draft, final form, under review
	Size/Scope of Knowledge Pool	Number and variety of users who have developed, implemented, tested, and refined products based on the standard
	Certification Status	The extent to which a certification program has been initiated or practiced to verify a product's conformance to standard
2. Market Conditions	Number of Product Sources	Number of contractors or developers offering products that follow a particular protocol standard
	User/Device Installation Base	Number of individuals or organizations who use the protocol products
	Product Stage of Development	Introduction, growth, maturity, or decline
	Contractor Commitment and Support	The extent to which contractors provide product support, implementation assistance, and other services to customers
3 Schedule	Product Readiness	The extent to which standards-related and manufacturing-related conditions affect product delivery
	Product Availability in 1991	The probability that a sufficient quantity and variety of products will be available in 1991
	Availability Date of Certified Products	The date(s) on which certified products are expected to be available
4. Cost	Conversion	Costs incurred to develop, test, and implement the target system
	Hardware and Software	e.g., network interface units, routers, protocol software, etc.
	New Operational Capabilities and Applications	Costs incurred to develop new APS operational capabilities and applications based on the target protocols
	Access to External Users and Data Resources	Costs incurred to implement standard file transfer, message handling, virtual terminal protocols
	Disposition of Equipment and Software	The extent to which previous equipment and software cannot be re-used
	Other	e.g., training, decreased productivity

- **Schedule:** The availability of protocol products in the form and times at which the PTO needs them:

  The PTO must meet APS operational, deployment, and acquisition schedules and budgets to meet the public's demand for technical information. Within the 1991- 1997 time period, the PTO expects to acquire a new APS network as well as additional workstations, storage devices, and applications. Following the alternative path should not adversely affect nor delay these activities.

- Costs: The costs expected to be incurred by each alternative in the years 1991 through 1997:

Conversion-related costs are one component of overall APS operating costs; future benefits from the alternative path chosen should compensate for the conversion costs incurred.

In evaluating the alternatives against the four criteria, in-house data communications specialists were consulted and current literature, product descriptions, and relevant MITRE documents were reviewed.

To evaluate the first two criteria, maturity of the protocol standards and market conditions, a marketing concept, the four-phased life cycle of product growth, was employed to determine a product's (or standard's) stage of development:

- **Introduction:** The product (or standard) is new so the product group or type, rather than the brand, is promoted; information and education are emphasized.

- **Growth:** Heavy competition occurs; marketing emphasizes one product's superiority over another's; persuasion, rather than information, is emphasized.

- **Maturity:** Competition continues; developers enhance or improve their products to differentiate them from other products or they concentrate their efforts on one or more particular market segments.

- **Decline:** Efforts are renewed to create demand for the product class, perhaps by identifying and catering to specialized needs [6].

In evaluating the third criterion, schedule, the analysis assumed that a 1991 conversion to either the DDN or OSI protocols could be completed before the scheduled deployments are initiated. Admittedly, several factors will dictate the actual duration, quality, and costs of a conversion; an examination of such factors, however, was beyond the scope of this analysis.

To evaluate costs, low, medium, and high relative cost evaluations were assigned to each cost factor based on the fact that there would be more devices, functional capabilities, and software to convert, replace, and dispose of in 1997 than in 1991 due to the acquisitions and deployments scheduled for the 1991-1997 period.

The analysis also distinguished constant costs from variable costs. Constant costs are those incurred within the 1991- 1997 period regardless of the chosen alternative (e.g.. the new network, additional workstations, routine maintenance, and user training). Variable costs are those incurred for items that are alternative-dependent (e.g., network interface units, protocol software, application conversions, and training of support personnel in the new protocols).

It was also assumed that any conversion from the XNS protocols would incur high costs because the customized programs that invoke the XNS functions would need to be rewritten or replaced. A 1997 DDN-to-OSI conversion, however, was expected to incur lower conversion costs because the software would have been appropriately designed before and during the XNS-to-DDN conversion to anticipate and provide for the requirements of the DDN-to-OSI conversion.

To allow PTO examiners to access external hosts that reside on OSI-based packet-switched networks, and to allow non-PTO users of the APS to access the APS from such facilities, this study assumed that the PTO would need to implement the following three OSI standard applications on the APS between 1991 and 1997: File Transfer, Access, and Management (FTAM), Consultative Committee for International Telegraph and Telephone (CCITT) Recommendation X.400 Message

Handling Systems (MHS), and the Virtual Terminal Protocol (VTP); the costs for such implementations would depend on the particular protocol suite in operation.

Finally, this study ignored the factors of quantity discounts that may apply to the purchases of hardware, software, and services (e.g., training) related to an alternative and, instead, estimated such costs in terms of unit costs such that the larger the number of units to be purchased, the higher the costs.

Evaluation of Alternatives

Table 2 summarizes the evaluation results of the three OSI conversion alternatives against the four criteria and their component factors.

XNS Alternative

The XNS Alternative is based on proprietary data communications standards that are some of the earliest LAN standards-the Xerox Corporation having designed and developed its "Experimental Ethernet" in 1975 [9]. While ethernet networks developed and implemented since then have continued to be popular, the proprietary XNS protocol standards have been replaced with Institute of Electrical and Electronics Engineers (IEEE) 802.3 LAN standards, the DDN protocols, and. to some extent, the OSI protocols.

This alternative is the most expensive alternative in terms of both the uncertainty regarding whether XNS-based products can be competitively acquired between 1991 and 1997 to meet APS requirements and schedules during that period, and the dollar costs of acquiring such products. Communication Machinery Corporation (CMC), the PTO's current supplier of XNS-based products, has dropped such products from its standard product line and is pursuing an OSI migration strategy that employs the DDN (or TCP/IP) protocols as an intermediate solution. Although CMC has indicated that it will continue to provide its XNS-based products to the PTO, the costs, duration, and degree of such support are unclear [2].

Large costs for modifying the custom software used to invoke XNS functions, as well as the applications that use it, are expected to be incurred from 199 1 to 1997, especially if the commercial products then available cannot operate with the XNS protocols or the APS-specific implementations of them. More software may also need to be developed to compensate for the limited functions that the XNS protocols perform. Such efforts will likely cause delays in the scheduling of APS implementations, tests, and deployments.

Finally, the likelihood that the PTO will not be able to re-use or "trade in" its XNS-based products when the OSI protocol conversion occurs in 1997 is an important cost consideration-in terms of costs to be avoided rather than costs to be incurred.

DDN Alternative

The DDN Alternative uses protocols that are based on more than 20 years of experience originating from the technology of the ARPANET, "the pioneer of packet switching networks," which was designed under a 1969 Defense Advanced Research Projects Agency (DARPA) research and development program [1]. The maturity of both the DDN protocol standards and the DDN product market, as well as its large installed base of users, attest to the protocol suite's utility and acceptance as a solution to a variety of problems. In fact, because of these conditions the Internet Activities Board (IAB), which governs the use of the DDN, has proposed that the major DDN protocols be made ANSI standards [8]. DDN protocol products are currently available from a large number of sources for a variety

Table 2. Conversion Alternative Evaluation Results

CRITERION	FACTORS	XNS	DDN	OSI
1. Maturity of the Protocol Standard	Status of the Protocol Standard	Specifications Complete	Errors Known, Identified	GOSIP based on Stable OSINET Agreements
	Size/Scope of Knowledge Pool	Small/Ethernet LAN	Large/PSNs, LANs, Internet	Medium/PSNs, LANs, Internet
	Certification Status	None	Ready	In Development
2. Market Conditions	Number of Product Sources	Small	Large	Medium
	User/Device Installation Base	Small	Large	Small
	Product Stage of Development	Decline	Maturity	Introduction
	Contractor Commitment and Support	None	Positive	Positive
3. Schedule	Product Readiness Factor(s)	Demand for XNS protocols, Contractor support (CMC)	Latest RFC, DDN Certification Requirements	GOSIP FIPS, OSI Certification Requirements
	1991 Product Availability (Prob)	Small	Large	Medium
	Expected Certification Year(s)	Not applicable	1989-1990	1991+
4. Cost	Conversion	High	High	Low-Medium
	Hardware and Software	High	High	Medium
	New OCs and Applications	High	Medium	Medium
	FTAM, MHS. VTP	High	Low	Low
	Disposition of Hardware/Software: XNS DDN	High	Low Medium	Low
	Other: Support Staff Trainrng	High (1997)	High (1991, 1997)	High (1991)

Key,

CMC: Communication Machinery Corporation
DDN. Defense Data Network
FIPS: Federal Information Processing Standard
FTAM File Transfer, Access, and Management
GOSIP Government OSI Profile
LAN Local Area Network
MHS, Message Handling System
OC Operational Capability
OSI: Open Systems Interconnection
OSINET: NIST/Industry PSN used for standards development
PSN. Packet Switched Network
RFC Request for Comments
VTP: Virtural Terminal Protocol
XNS, Xerox Network Systems

Table 3. Evaluation Summary

CRITERION	XNS	DDN	OSI
Maturity of the Protocol Standard	Decline	Maturity	Growth
Market Conditions	Decline	Maturity	Introduction
1991 Product Availability (Prob)/ Expected Certification Year(s)	Small/ N/A	Large/ 1989-90	Medium/ 1991+
Costs	High	Medium	Low-Medium

of networks, operating systems, and applications; such availability is expected to continue into the 1990s.

Since the GOSIP was formally approved as a FIPS in August 1988, the DOD has adopted the OSI protocols as full co-standards and has initiated transition and interoperability activities [5]. In response to this conversion effort, the DDN-based products available in the 1989- 1997 period are expected to incorporate or facilitate conversion to the OSI protocols: they may support both the DDN and OSI protocol suites or allow DDN-based protocol components to be replaced with OSI-based ones at less-than-full-product cost.

The DDN Alternative has medium overall costs plus opportunities for cost savings. DDN protocol products, if wisely selected, may be usable with the OSI-based APS, thus reducing hardware, software, and DDN protocol disposal costs.

OSI Alternative

In 1977, recognizing the difficulty of interconnecting diverse devices among government, academic, and commercial data communications users, the IS0 established committees to develop a common architecture with which such communications could be provided. In 1984, the IS0 published the "Open Systems Interconnection-Basic Reference Model" (IS 7498) which distributed data communications functions among seven OSI layers. The CCITT, pursuing a similar objective. produced its Recommendation X.200, "Reference Model of Open Systems Interconnection" [1]. Standardization efforts pursued since then by IS0 and other standards groups have resulted in the OSI standards currently in effect.

OSI protocol standards are at most seven years old, and OSI-based products are beginning to be developed and marketed. While OSI supports more functions than the DDN and XNS protocol suites, a limited number of protocols and applications (e.g., FTAM, CCITT X.400) have been developed into products. The number of OSI product users and sources is expected to increase as the OSI protocols mature, as GOSIP-certified products are implemented, and as additional protocols are added to the GOSIP. Such maturity of standards and products, however, is not expected to occur when the PTO must decide, plan, and initiate its implementation. Since only one protocol conversion in 199 1 is required and a small quantity of XNS-based hardware and software will need to be replaced, the OSI Alternative has low to medium costs.

Recommendations

Based on the four criteria and their order of importance (see table 3), MITRE recommended that the PTO follow the DDN Alternative-that is, convert to the DDN protocols first, then convert to the OSI protocols in 1997. With the XNS Alternative, the APS will be locked into an obsolete standard, data communications capabilities will be limited to those defined in the XNS protocols or provided in the customized software that invokes them, and accommodating new functions, capabilities, and requirements will incur very high costs.

Mature DDN protocol standards and competitive market conditions indicate the continued popularity, cost effectiveness, and applicability of the DDN protocols to government and other users' requirements. Market projections for the 1988-to- 1993 period indicate that both DOD and civilian federal government agencies will continue acquiring and implementing DDN protocols; by one estimate, government sales of DDN protocol products and services are expected to drop gradually from 1988's 60 percent share to 48 percent of the market in 1993 [7]. In addition, several protocol vendors' open systems strategies call for supporting the DDN protocols as a near-term solution and migration path to the OSI protocols. Contractor commitment and assistance, therefore, should also be readily available.

Admittedly, following the OSI Alternative would eliminate a 1997 conversion and avoid related costs. However, based on the three higher ranked criteria-maturity of the protocol standard, market conditions, and schedule-a 1991 conversion of the APS to the OSI protocols would be premature. However, a 1993 or 1994 conversion to the OSI protocols might be a more practical and less costly solution that allows for the maturing of both the OSI standards and the products based on them. While the PTO's plans and schedule, which only allow for a protocol conversion either in 1991 or in 1997, preclude the PTO's pursuit of this "fourth alternative," this solution will likely apply to many government agencies whose missions allow it.

The DDN Alternative is only a high-level description of the conversion path that the PTO will follow in making the APS an open system. Clearly, the PTO will also need to perform more detailed analyses to determine the APS's specific protocol requirements, to justify delaying a GOSIP-based implementation of the APS, to determine the acquisition strategy with which the PTO will specify, purchase, implement, and dispose of its DDN-based products, and to develop the plans with which the two conversions will be performed.

References

1. Black, Uyless, 1989, *Data Networks: Concepts, Theory, and Practice,* Englewood Cliffs, NJ: Prentice-Hall, Inc.

2. Foreman, Elizabeth A., 3 1 January 1989, "Communication Machinery Corporation's (CMC) Implementation Plans for the Open Systems Interconnection (OSI) and Xerox Network Systems (XNS) Protocol Suites," MITRE Technical Letter (W21-TL-25) to Dave Stewart, U.S. Patent and Trademark Office.

3. Foreman, Elizabeth A., March 1990, *Automated Patent System Open Systems Interconnection Conversion Analysis,* MITRE Working Paper, WP-89WOO477, The MITRE Corporation, McLean, VA.

4. Foreman, Elizabeth A., March 1990, *Automated Patent System Open Systems Interconnection Conversion Strategy Analysis,* MITRE Technical Report MTR-89W00207, The MITRE Corporation, McLean, VA.

5. Gross, Phillip, and Richard Wilder, May 1988, *The Department of Defense Open System interconnection (OSI) implementation Strategy*. MITRE Technical Report. MTR-87W00155, The MITRE Corporation, McLean, VA.

6. Marcus, Burton H., et al., 1975, *Modern Marketing,* New York, NY: Random House, Inc., pp. 294-295.

7. Masud, S. A., 7 August 1989, "TCP/IP Seen Staying Despite GOSIP Invasion," *Government Computer News*, p. 25.

8. Tarquinio, Sally W., 25 February 199 1, The MITRE Corporation, telephone conversation regarding potential ANSI standardization of DDN protocols.

9. Xerox Corporation, Digital Equipment Corporation, Intel Corporation, 30 September 1980, *The Ethernet, A Local Area Network: Data Link Layer and Physical Layer Specifications.*

About the Author

Elizabeth (Beth) Foreman is a member of the technical staff of the Advanced Information Systems Division. She holds a B.A. in Mathematics and an M.B.A. in information Systems Technology. Beth is currently assisting the U.S. Patent and Trademark Office in defining the near- and long-term configurations for its Automated Patent System (APS) and in determining the problems, factors, and strategies involved in making the APS an open system. Beth's professional interests include telecommunications network architectures and protocols; applications portability and open system environments; protocol conversion analysis and planning; and systems analysis, design, and testing.

Telecommunications Pricing as a Vector Space Optimization Problem

Thomas B. Fowler

A portion of the general problem of pricing telecommunications services in the fairest and most competitive manner is posed as a problem of finding the optimal mapping from a very high-order vector space (with dimension equal to the maximum number of point-to-point connections) to a much lower order space, that of a small number of price formulae. The objective is to make the price of the calls between any two physical locations as calculated in the lower order space as close as possible to the actual cost as given in the high-order space, in the sense of minimizing an error function. This optimal mapping is dependent upon both the call distribution profiles and the call price versus time functions. Based on this formalism, it can be shown that there exist calls which will be priced higher with the optimal mapping than with the other, suboptimal mappings.

Introduction

Telecommunications systems involve traffic between a large number of service locations. A separate price formula for each possible connection may be assumed to exist, which gives call price as a function of call duration. In general, however, it is not practical to use this large set of formulae to bill customers because of the extreme complexity such a scheme would entail, and the difficulty the customers would have when trying to estimate their telecommunications costs. Therefore, vendors typically use a small set of *m* price formulae or tariffs to price all calls. These tariffs are frequently based on distance, with calls over greater distances costing more per unit time. The general problem addressed here is how to map the large number of connections and their associated price formulae into the smaller number of tariffs actually used to charge customers.

The problem may be posed as a vector space optimization problem, with the original price formulae as vectors in a higher order vector space and the smaller set of *m* tariffs as vectors in a lower order space. The objective is then to make the price of a call between any two physical locations, as charged to the customer using one of the *m* tariffs (vectors) from the lower order space, as close as possible to the price as it would be calculated using the appropriate formula (vector) from the higher order space. That is, one wishes to minimize the distance between the vector from the higher order space and the vector from the lower order space. If the price formula from the higher order space is assumed to reflect actual costs of service delivery, such a mapping will be optimal for both vendors and consumers

1 The new Federal Telecommunications System 2000 (TSD2000) is priced in this manner, for example, as are most residential and commercial long-distance voice calls FTS2000 also bases its pricing structure on volume "bands."

in the sense that all calls are priced in the most fair manner, i.e., each pays its own way (or closer to it on average) than with any other allocation system. Such an optimal mapping will be a function of the call distribution profile as well as the original price functions. Based on this formalism, it can readily be shown that there exist sets of calls which will be priced higher with the optimal mapping than with other, suboptimal mappings. That is, a vendor can undercut an optimally fair vendor's prices for some traffic, although he will then have to charge more than the optimally fair vendor's prices for other traffic to compensate (assuming total revenue is to be kept constant).

Assumptions

The given telecommunications system has n service locations with bi-directional connections, yielding $q = \frac{n(n-1)}{2}$ total possible connections, numbered $c_1, c_2, \ldots c_q$. Traffic distribution as a function of time (number of calls of a given length) for each c_i is known and designated by $\mu_i(t)$, where t is the call length. The desired pricing function (generally that needed to recover network operating cost) is also known for each c_i, and is given by a function $\gamma_i(t)$, where t is the call length. Total revenue r_i associated with connection c_i is given by the Riemann-Stieltjes integral

$$r_i = \int_0^{t_{max}} \gamma_i(t) d\mu_i(t) \tag{1}$$

in which the price charged $\gamma_i(t)$ for a call of length t is appropriately weighted by the volume of traffic of that length $\mu_i(t)$. The $\mu_i(t)$ functions are assumed to be of bounded variation to guarantee convergence of the integral (1); for practical cases, of course, this would not be a problem.

Optimization Problem Formalism

The q original price functions $\gamma_i(t)$may be taken as vectors in a q-dimensional vector space H_1 with inner product

$$<\gamma_1, \gamma_2> \;=\; \int_0^{t_{max}} |\gamma_1\gamma_2| dt \tag{2}$$

The m price functions corresponding to the reduced number of price bands also form a Hilbert space H_2. These functions will be designated $\xi_1(t)$, $\xi_2(t), \ldots \xi_m(t)$. They may be assumed to have same inner product as H_1. H_2 is not necessarily a subspace of H_1; it will be a subspace only if each $\xi_i(t)$ is expressible as a linear combination of $\gamma_j(t)$ for all j.

The optimal mapping is a functional M^* from the set of price versus time functions $\gamma_i(t)$ in H_1 to those in H_2, the $\xi_j(t)$ functions. M^* maps each γ_i into one and only one ξ_j. The object is to find the optimal mapping $M^*: \{\gamma_i(t) \rightarrow \xi_j(t)\}$ such that a certain error or penalty function η is minimized.

A quadratic objective function (actually an integral functional) may be chosen as

$$\eta^2 = \sum_{i=1}^{q} \int_0^{t_{max}} [\xi_\alpha - \gamma_i]^2 d\mu_i \tag{3}$$

where $\alpha \in \{1..m\}$ and the time argument has been suppressed. The α's are chosen to minimize η^2. This functional has the significant advantage of preserving revenue over any given time period, as will be shown below.

Nature of Optimization Problem

This is not a standard projection theorem optimization problem [1] for two reasons:

(1) The optimal subspace is not given, only its dimension (=m). In this respect, the problem is similar to the well-known model-reduction problem of large-scale system theory, in which the underlying systems are nonlinear [2]. (2) The error functional η^2 involves integrals different than those of the inner product for H_1 and H_2. This means that if η is taken as a measure of distance between the original price function and the reduced-order price function, it corresponds to a different distance measure than that used in the original spaces.

The solution methodology is as follows: divide the time interval $0 \le t \le t_{max}$ into a large number of subintervals and find the optimal solution for each subinterval in the sense of minimizing the error functional over that subinterval; the overall solution, consisting of the solution pieces for each subinterval, will then be globally optimal. This can be readily accomplished by taking the subintervals sufficiently small so that over each, the $\mu_i(t)$, $\gamma_i(t)$, and $\xi_j(t)$ functions can be considered approximately constant, and then solving pointwise for the exact solution. The method is formalized in the following theorem:

Theorem 1. The solution to the optimization problem can be found and is given by

$$\xi^*(t) = \frac{\sum_{i=1}^{q} \gamma_i(t)\mu_i(t)}{\sum_{i=1}^{q} \mu_i(t)} \tag{4}$$

for the case of m=1 and for the case of m>1 by

$$\xi_j^*(t) = \frac{\sum_{i=1}^{q^{(j)}} \gamma_i^{(j)}(t)\mu_i^{(j)}(t)}{\sum_{i=1}^{q^{(j)}} \mu_i^{(j)}(t)} \tag{5}$$

where the superscript (j) designates those $\gamma_i(t)$ vectors mapped to $\xi_j(t)$ and $q^{(j)}$ their total number.

Proof. The proof is a slight variation on a theorem of Luenberger [3], which states that a necessary condition for a real-valued functional f to have an extremum at a point ξ^* is for its Gateaux differential to be zero at that point, i.e.,

$$\delta f(\xi^*;h) = \frac{df}{d\xi}h = 0 \quad \forall h \quad \Rightarrow \quad \frac{df}{d\xi} = 0 \tag{6}$$

Begin with case m=1. Partition interval $[0 \ldots t_{max}]$ into p segments with midpoint t_k and length Δt_k. Observe that

$$\int_0^{t_{max}} f(t)d\mu_i = \lim_{p \to \infty} \sum_{k=1}^{p} f(t_k)\mu_i(t_k)\Delta t_k \tag{7}$$

Applying this approximation to the calculation of η yields

$$\eta^2 = \sum_{i=1}^{q} \int_0^{t_{max}} [\xi(t) - \gamma_i(t)]^2 d\mu_i \approx \sum_{i=1}^{q}\left(\sum_{k=1}^{p} [\xi(t_k) - \gamma_i(t_k)]^2 \mu_i(t_k)\Delta t_k\right) \tag{8}$$

Reversing the order of summation,

$$\eta^2 \approx \sum_{k=1}^{p}\left(\sum_{i=1}^{q} [\xi(t_k) - \gamma_i(t_k)]^2 \mu_i(t_k)\Delta t_k\right) = \sum_{k=1}^{p} \eta_k^2 \tag{9}$$

$$\eta_k^2$$

The optimization can then be carried out term-by-term on the η^2_k terms:

$$\frac{d\eta_k^2}{d\xi_k} = \frac{d}{d\xi_k}\left(\sum_{i=1}^{q} [\xi_k - \gamma_i(t_k)]^2 \mu_i(t_k)\Delta t_k\right) = \sum_{i=1}^{q} \frac{d}{d\xi_k}([\xi_k - \gamma_i(t_k)]^2 \mu_i(t_k)\Delta t_k) \tag{10}$$

where $\xi_k = \xi(t_k)$. Carry out the differentiation and regroup the terms:

$$\frac{d\eta_k}{d\xi_k} = \frac{1}{2\eta_k}\left(\sum_{i=1}^{q} 2[\xi_k - \gamma_i(t_k)]\mu_i(t_k)\Delta t_k \right)$$

$$= \frac{1}{\eta_k}\left(\xi_k \sum_{i=1}^{q} \mu_i(t_k)\Delta t_k - \sum_{i=1}^{q} \gamma_i(t_k)\mu_i(t_k)\Delta t_k \right) \tag{11}$$

Set the right-hand side equal to 0 and solve for ξ_k,

$$\xi_k = \frac{\sum_{i=1}^{q} \gamma_i(t_k)\mu_i(t_k)}{\sum_{i=1}^{q} \mu_i(t_k)} \tag{12}$$

This solution is just the centroid of the i values of $\gamma_i(t_k), \mu_i(t_k)$, and it must be globally optimum since there is only one centroid. The solution construction method implies that the exact solution can be found pointwise to any desired degree of accuracy by taking sufficiently many points k.

The solution methodology in the case of $m>1$ is slightly more complex in that it requires enumeration of all possible mappings of the γ_i functions to the ξ_j functions, and then separately optimizing each to determine the one with the smallest value of η^2.

Starting with the second part of this procedure, assume that a mapping is given and, as indicated above, let $\gamma_i^{(j)}$ denote those γ_i functions associated with reduced-order price function ξ_j. Then

$$\eta^2 = \sum_{j=1}^{m} \sum_{i=1}^{q^{(j)}} \int_0^{t_{max}} [\xi_j(t) - \gamma_i^{(j)}(t)]^2 d\mu_i^{(j)}(t) \tag{13}$$

Clearly the inner sums can be optimized separately since the $\xi j(t)$ functions are independent, which implies that the foregoing method can be used on the inner sums separately to compute each optimal $\xi j^*(t)$, yielding (5).

The overall solution is obtained in the following way: (1) enumerate all possible mappings $Mi: \{\gamma i\} \rightarrow \{\xi j\}$, where each γi is mapped to a single ξj; (2) for each Mi, calculate optimal solutions $\{\xi j(t)\}$ and corresponding ηi, as described above; (3) choose mapping M^* with smallest $\eta 2$.

QED

Theorem 2. The optimal reduced-order price functions preserve total revenue and total revenue distribution.

Proof. Considering again the case of $m=1$, note that for any given time interval Δt, total revenue for the original set of price functions is given by

$$r_k = \sum_{i=1}^{q} \gamma_i(t_k)\mu_i(t_k)\Delta t_k \tag{14}$$

For the reduced-order price function, revenue is given by

$$\sum_{i=1}^{q} \xi_k \mu_i(t_k)\Delta t_k = \xi_k\left(\sum_{i=1}^{q} \mu_i(t_k) \right)\Delta t_k$$

$$= \frac{\sum_{i=1}^{q} \gamma_i(t_k)\mu_i(t_k)}{\sum_{i=1}^{q} \mu_i(t_k)}\left(\sum_{i=1}^{q} \mu_i(t_k) \right)\Delta t_k$$

$$= \sum_{i=1}^{q} \gamma_i(t_k)\mu_i(t_k)\Delta t_k$$

$$= r_k \tag{15}$$

For the case of $m>1$, the argument proceeds in the same manner except that the sums are taken over each reduced-order price function $\xi_j(t)$ and its associated original price functions $\gamma_i^{(j)}(t)$ separately.

QED

Case of Given Reduced-Order Function x(*t*)

The basic method can also be used to calculate the optimal constant coefficient α needed to multiply some given function $\xi(t)$, such as a flat rate price, so as to minimize η, given as

$$\eta^2 = \sum_{i=1}^{q} \int_0^{t_{max}} [\alpha\xi - \gamma_i]^2 d\mu_i \tag{16}$$

In this case, pointwise optimization cannot be done and global optimization over the entire time interval $[0\ldots t_{max}]$ is required. Let $f = \frac{d\eta}{d\alpha^*}$. Then a minimum (perhaps local) occurs for $f=0$. The optimal α is obtained from the Newton-Raphson recursion

$$\begin{aligned} \alpha(n+1) &= \alpha(n) - [df/d\alpha]^{-1} \cdot f \\ &= \alpha(n) - [\partial^2\eta/\partial\alpha^2]^{-1} \cdot f \end{aligned} \tag{17}$$

Revenue will not in general be conserved, however. If an optimal linear combination of several given reduced-order price functions should be desired, the method can readily be generalized to handle such cases by making the unknown constants form a vector A and utilizing the matrix version of the Newton-Raphson iteration algorithm. However, details of that generalization will not be discussed here.

Example

Assume there are three service locations in some simple telecommunication system. This implies three connections (q=3), taken to have the following price functions:

$$\begin{aligned} &c_1 \text{ (node 1 - node 2):} \quad \gamma_1(t) = t/2 \\ &c_2 \text{ (node 1 - node 3):} \quad \gamma_2(t) = 1 - e^{-t} \\ &c_3 \text{ (node 2 - node 3):} \quad \gamma_3 = \begin{cases} t & t \le 2 \\ 2 & t > 2 \end{cases} \end{aligned} \tag{18}$$

The corresponding traffic distributions (measures) are assumed to be:

$$\begin{aligned} &c_1 \text{ (node 1 - node 2):} \quad \mu_1(t) = e^{-t} \\ &c_2 \text{ (node 1 - node 3):} \quad \mu_2(t) = \beta(t), \\ &\qquad\qquad\qquad\qquad\qquad p = 2, \quad q = 3 \\ &c_3 \text{ (node 2 - node 3):} \quad \mu_3(t) = 0.1 \end{aligned} \tag{19}$$

The objective is to find the optimal reduced-order price function $\xi(t)$ such that

$$\eta^2 = \sum_{i=1}^{3} \int_0^{t_{max}} [\xi(t) - \gamma_i(t)]^2 d\mu_i(t) \tag{20}$$

is minimized. The optimal solution function $\xi^*(t)$ is illustrated in figure 1, and the corresponding total revenue distribution curve in figure 2. The total revenue generated is 3.235 units. Figures 3 through 5 show the revenue distribution curves for the original three $\gamma_i(t)$ functions, together with the revenue corresponding to the same $\mu_i(t)$ but with the new single optimal price function $\xi^*(t)$. The same set of curves for the case of the optimal flat-rate function, calculated using (17), is shown in figures 6 through 10. The optimal flat rate in this case is 1.07353; the total revenue generated is 3.310 units, slightly more than that with the optimal curve.

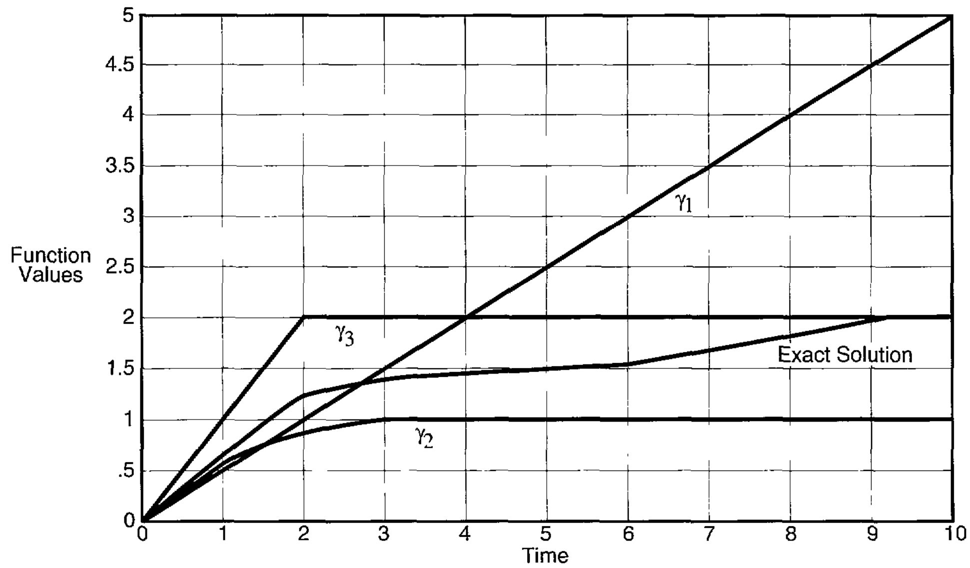


Figure 1. Optimal Single Reduced-Order Price Function $\xi(t)$ for Example Discussed in Text
This function gives the call price, as a function of time, which will be charged uniformly for calls placed over all three of the original connections. Also shown, for comparison, are the original three price functions $\gamma_i(t)$.

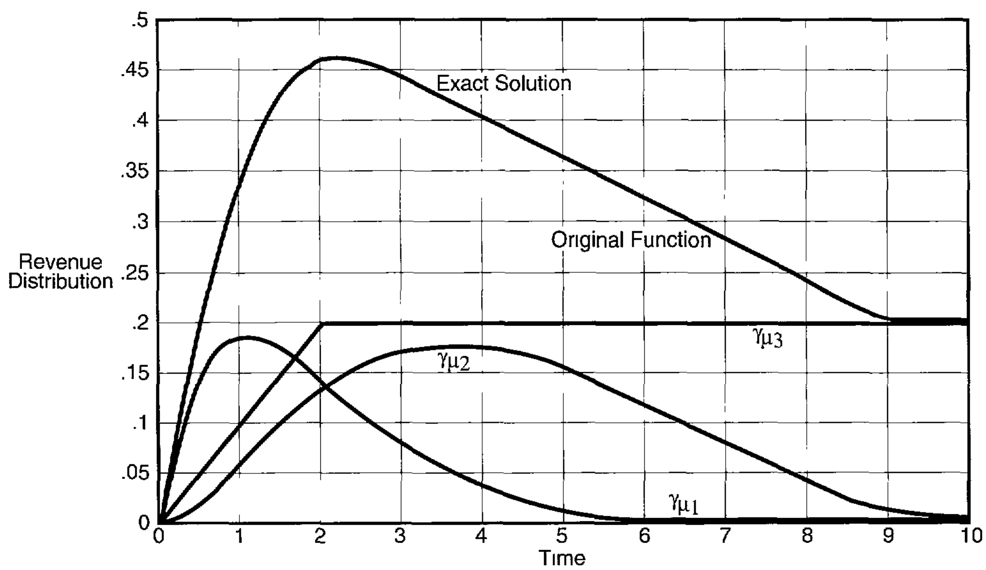


Figure 2. Revenue Distribution as a Function of Call Length Generated by Reduced-Order Price Function $\xi(t)$
This curve coincides with the sum of the revenue distribution curves under the original price functions, as shown. The total money collected by the vendor is the area under this curve. The money collected for calls of lengths between t_1 and t_2 is the area of the curve between those two abscissa points. The individual revenue distribution curves are also included for comparison purposes.

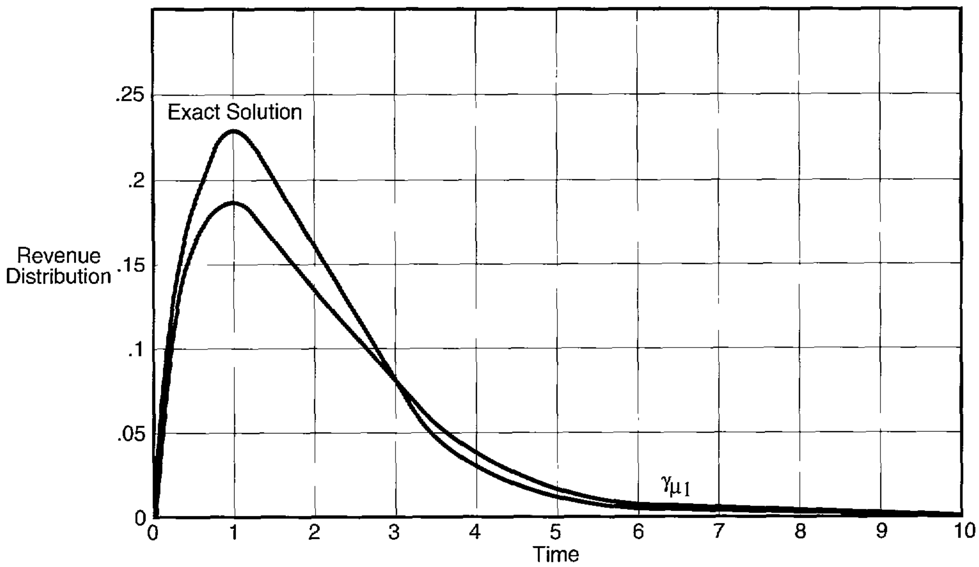


Figure 3. Revenue Distribution Curves for the Original Price Functions $\gamma_1(t)$ and $\mu_1(t)$, together with the Revenue Distribution Curve Corresponding to $\xi(t)$ and $\mu_1(t)$

The revenue distribution curve gives the revenue distribution as a function of call length that the communications vendor would collect using the single reduced-order price function. As before, the revenue collected corresponds to the area under the curve.

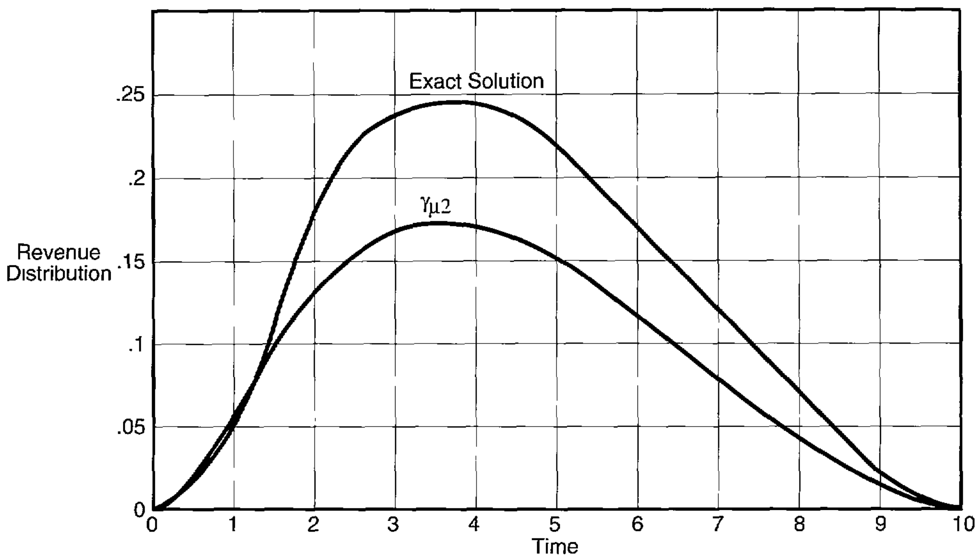


Figure 4. Revenue Distribution Curves for the Original Price Functions $\gamma_2(t)$ and $\mu_2(t)$, together with the Revenue Distribution Curve Corresponding to $\xi(t)$ and $\mu_2(t)$

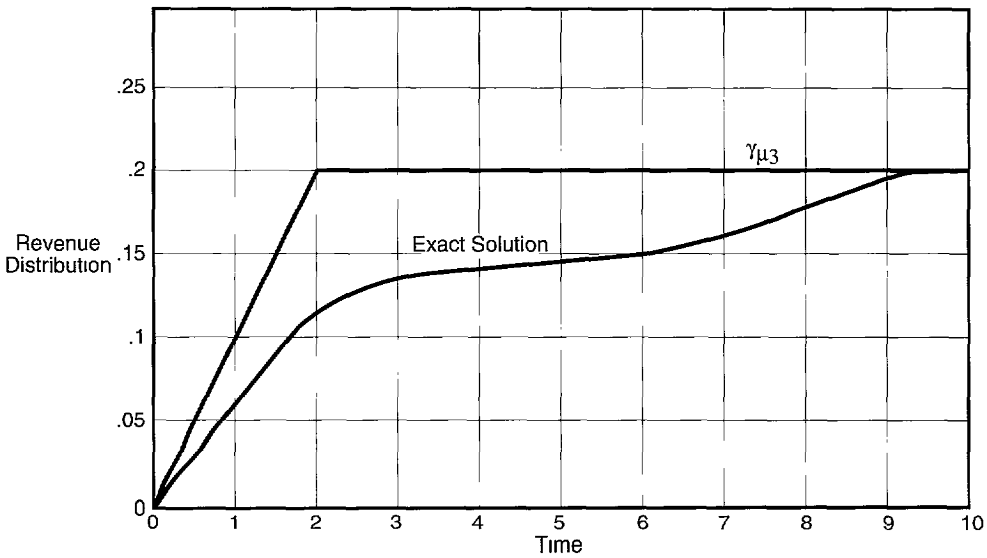


Figure 5. Revenue Distribution Curves for the Original Price Functions $\gamma_3(t)$ and $\mu_3(t)$, together with the Revenue Distribution Curve Corresponding to $\xi(t)$ and $\mu_3(t)$

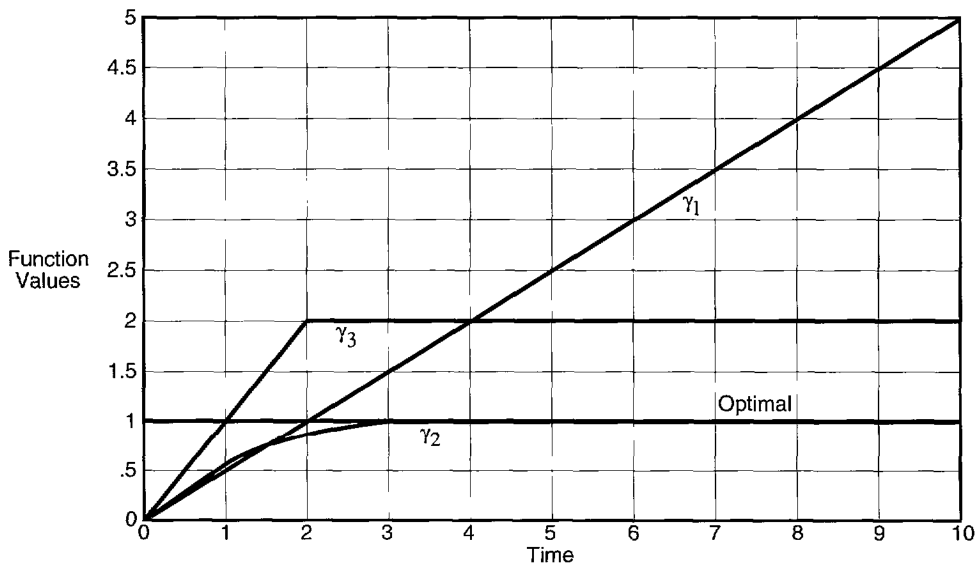


Figure 6. Optimal Flat-Rate Price Function for Example Discussed in Text
This function gives the flat call price which will be charged uniformly for calls placed over all three of the original connections.

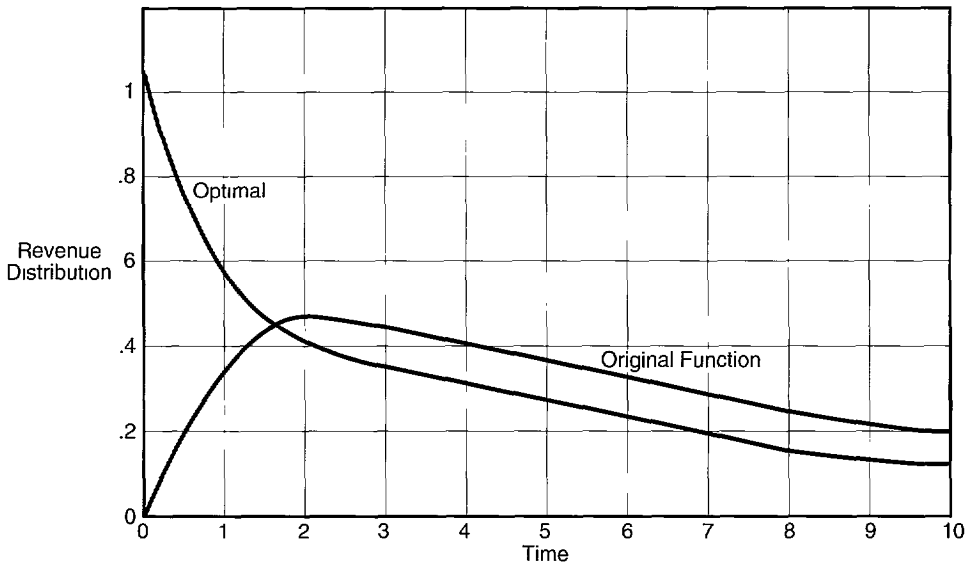


Figure 7. ***Total Revenue as a Function of Call Length Generated by Optimal Flat-Rate Price Function***
Also shown is the sum of the revenue curves under the original price functions.

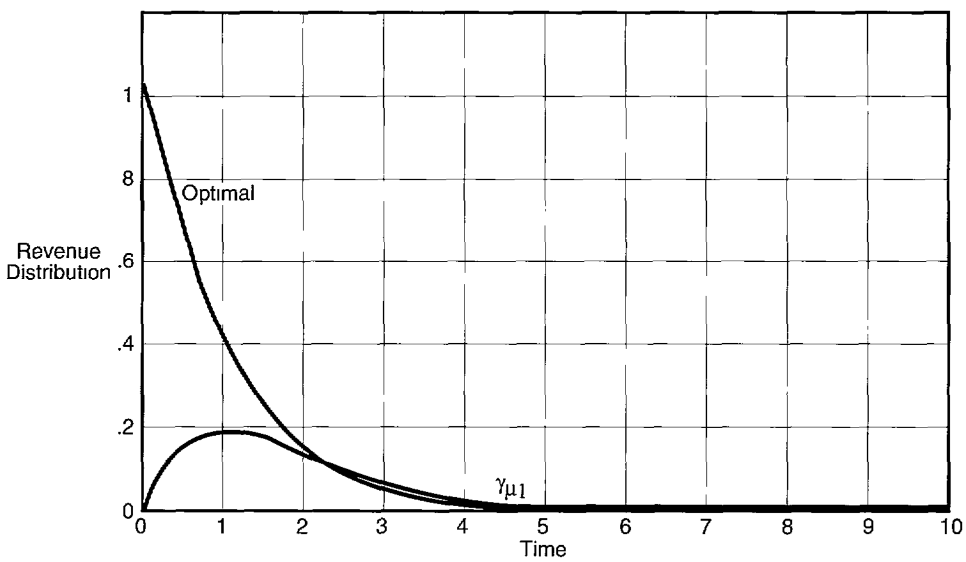


Figure 8. Revenue Distribution Curves for the Original Price Functions $\gamma_1(t)$ and $\mu_1(t)$, together with the Revenue Distribution Curve Corresponding to the Optimal Flat-Rate Price and $\mu_1(t)$

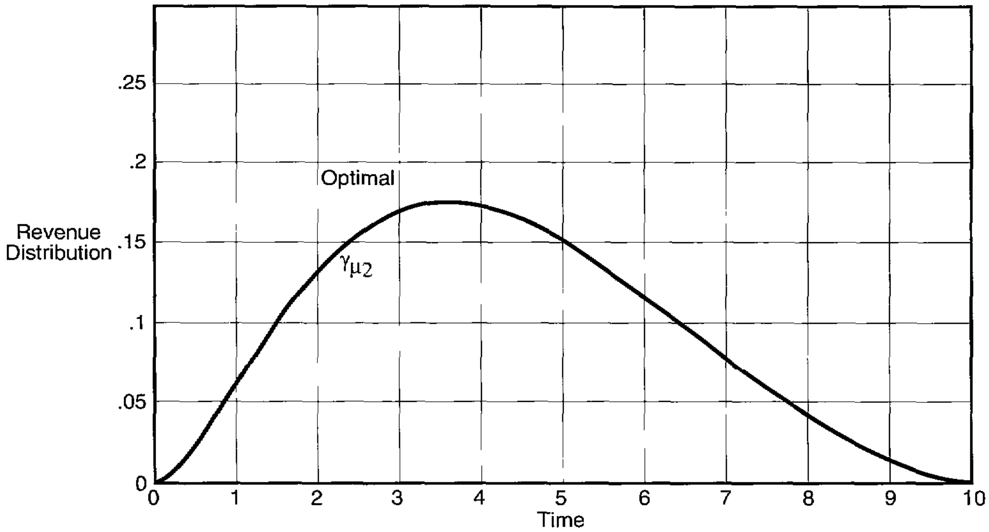


Figure 9. Revenue Distribution Curves for the Original Price Functions $\gamma_2(t)$ and $\mu_2(t)$, together with the Revenue Distribution Curve Corresponding to the Optimal Flat-Rate Price and $\mu_2(t)$

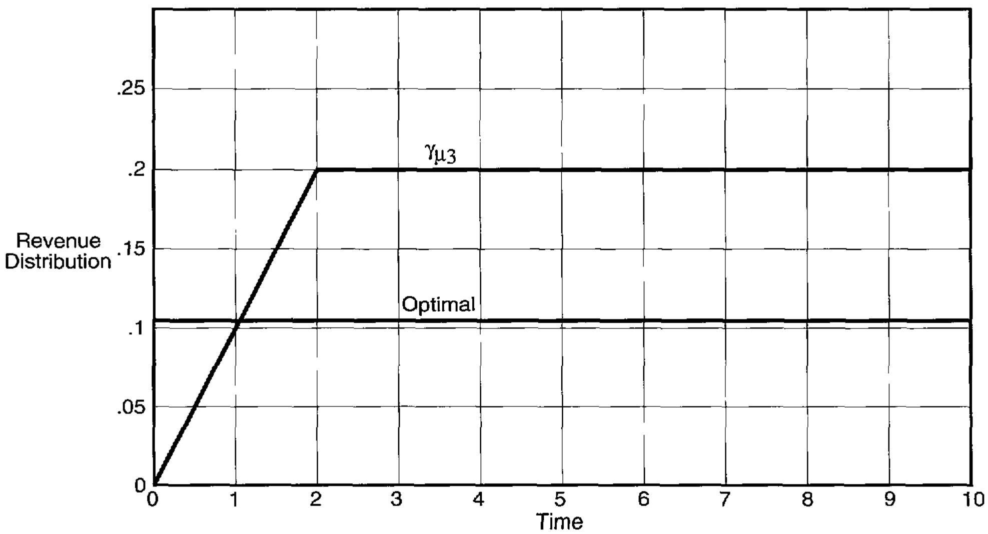


Figure 10. Revenue Distribution Curves for the Original Price Functions $\gamma_3(t)$ and $\mu_3(t)$, together with the Revenue Distribution Curve Corresponding to the Optimal Flat-Rate Price and $\mu_3(t)$

Existence of Mappings M:$\{\gamma_i(t) \rightarrow \mu_j(t)\}$ for which at Least One Call is Cheaper than in Optimal Mapping M^*

The existence of such mappings is relevant to the functioning of certain price control mechanisms, such as those in the new Federal Telecommunications System 2000, FTS2000 [4]. One of these mechanisms, as originally implemented, called for "capping" any FTS2000 call price at a value corresponding to its price on a publicly available network. This mechanism was not perceived as fair by the FTS2000 vendors because it penalized them for calls priced higher than those of the public domain, but provided no compensating adjustment for calls priced *lower* in FTS2000. To understand why such a situation is virtually certain to exist, pick some desired call length t^k. From the set of optimal reduced-order functions $\Xi^*=\{\xi_1(t)\ldots\xi_m(t)\}$, find some pair $\{\xi_i(t),\gamma_j(t)\}$ such that $\xi_i(t)$ maps to $\gamma_i(t)$ and $\xi_i(t_k)>\gamma_j(t_k)$. The existence of this pair is not guaranteed, but the probability that it does not is vanishingly small for even low numbers of locations n. (Such a pair would fail to exist only in the case that for the given t_k, the reduced-order price functions $\xi_i(t)$ happened to have the same value as their corresponding $\gamma_j(t)$ functions. If this occurs, choose a different t_k; such a point must exist since the $\gamma_j(t)$ are assumed independent.) A non-optimal set of reduced-order functions can be created by replacing whichever $\xi_i(t)$ maps to $\gamma_j(t)$ by $\gamma_j(t)$ itself and supplying appropriate constants to make total revenue equal to original revenue. The new set of reduced-order functions $\{a_1\xi_1(t),\ldots\gamma_j(t)\ldots a_m\xi_m(t)\}$, will be such that $\eta_{new\,set} > \eta^*$ but the cost for calls on connection c_k of length t_k will be less than under the optimal set.

Use of Methodology

Utilizing this methodology, the degree of optimality of current telecommunications pricing schemes can be analyzed, and recommendations made as to the degree of fairness of these schemes. In addition, the method can be used to derive new pricing schemes, and can be applied to any system in which services are rendered to public or private clients, including telecommunications in all its forms (electronic mail, etc.). A possible application may be in the area of telecommunications proposal preparation and evaluation: vendors could be required to structure their price tables in such a way that the error η is within certain bounds, for example.

References

1. Luenberger, D. G., 1969, *Optimization by Vector Space Methods,* Wiley & Sons, New York, pp. 49ff.

2. Fowler, T. B., August 1988, "A Time-Domain Solution Approach to Model Reduction," *IEEE Transactions on Circuits and Systems*, Vol. 35, No. 8, pp. 1020-1024.

3. Luenberger, op. cit., Section 7.4, Theorem 1, pp. 178.

4. General Services Administration, September 1988, *FTS2000 Contracts*, as amended, Section H.10, Washington, DC.

About the Author

Thomas Fowler received his B.A. in Philosophy and his B.S. in Electrical Engineering from the University of Maryland in 1969 and 1972, respectively. He also holds an MS. degree in Electrical Engineering from Columbia University and an Sc.D. degree from George Washington University. Tom joined MITRE in 1973 as a member of the technical staff first working on the TOCCOT project and then on Air Traffic Control projects. From 1982 to 1987, he was on the faculty of Christendom College in Front Royal, Virginia, and did consulting work. Tom rejoined the technical staff at MITRE in 1987, in the Advanced Information Systems Division where he is currently a principal engineer. He also serves as chairman of the Nonlinear Dynamics and Chaos cluster group and teaches graduate courses in system and control theory at George Washington University. Tom is interested in nonlinear and large-scale systems, especially as applied to biological and social problems.

GSA Government Open Systems Interconnection Profile (GOSIP) Registration Implementation

Victor J. Hsin
Frank E. Ferrante

The U.S. Government Open Systems Interconnection Profile (GOSIP) Version 1.0 became a compulsory and binding specification for the U.S. government in all of its telecommunications network and computer system solicitations on 15 August 1990. In order to identify the resources within GOSIP environments, a centralized registration is necessary. The National Institute of Standards and Technology (NIST) has delegated this operational registration responsibility to the General Services Administration (GSA). This paper describes the GOSIP registration procedures as currently operating in the U.S. government arena.

When considering the communications among different vendors' systems, a set of common communications standards is needed. Traditionally, vendors have locked users into proprietary solutions by designing architectures and protocols which are functionally incompatible with other vendors' systems. To enable communications between systems, proprietary gateway systems have been required to perform protocol and data format conversion functions. To address the interoperability problem among vendors, standards bodies have developed a common architecture called the Open Systems Interconnection (OSI) model. Government organizations all over the world, as very large buyers of information technology products, technology, and telecommunications services, have one of the clearest mandates for backing the internationally accepted set of standards being promulgated under the OSI banner. In the United States (U.S.), this backing comes in the form of the U.S. Government OSI Profile, or U.S. GOSIP. U.S. GOSIP is similar in purpose and content to the United Kingdom GOSIP, the Swedish GOSIP, and the OSI profiles of other governments (e.g., Canada and Australia). The U.S. GOSIP, however, is the only one of the several that have been defined which is mandated. The others show their commitment to the international OSI standards; however, they serve primarily as references and are not considered mandatory [1].

The U.S. GOSIP became a compulsory and binding specification for the government's telecommunications and computer solicitations on 15 August 1990.

Importance of Registration

In order to communicate with other users or hosts, it is necessary to identify the objects involved in communication. These objects have names and addresses. A name is a collection of attributes that identify an object within a domain. An address specifies the location of an object. The names and address must be registered with a registration authority so that they will not be used for more than one object, i.e., names and addresses must be unambiguous [2].

Without registration authorities, chaos will result with random name and address values being assigned to objects.

Registration authorities are created to register names of objects and, in some cases, to advertise these names. For example, the telephone companies assign numbers to subscribers and publish some of the numbers in a telephone directory. Registration authorities and procedures are essential in the OSI environment. Without registration authorities, chaos will result with random name and address values being assigned to objects. Since systems would not be able to identify themselves uniquely, the orderly communication among systems would become infeasible. In the OSI world, some names are included and registered in the standards. If users want to exchange objects not included in the standards, they must be registered somewhere so that no one else will use the same number for a different object or a different number for the same object. There are a number of possibilities for assigning object identifiers. We will describe several that are particularly relevant to GOSIP after reviewing U.S. GOSIP status.

U.S. GOSIP Status

U.S. GOSIP provides the specification for the U.S. federal government-approved applications of OSI. It is a user profile which is defined according to the needs of a community of interest and creates an opportunity for each federal agency to assert control over future procurements. U.S. GOSIP is very much a product of national consensus, since it is based on the OSI Implementors Workshop (OIW) Stable Implementation Agreements [3,4], and is authored by working groups comprised of representatives from many U.S. government agencies. The challenge facing both vendors and customers is how to implement OSI standards economically in products within the shortest period of time. Open adoption of GOSIP as a long-term strategic initiative is envisioned to lead to evolution of current systems into a GOSIP-compliant, interoperable set of computer systems within agencies.

The National Institute of Standards and Technology (NIST) has made a determined effort to speed the release of products by publishing GOSIP as a Federal Information Processing Standard (FIPS 146) [5], which took effect on 15 February 1989. An 18-month grace period was allowed during which time the General Services Administration (GSA) and other federal agencies were permitted to acquire systems using alternative protocols which provided equivalent functionality to the GOSIP protocols.

In U.S. GOSIP implementation, NIST has delegated operational registration responsibility to the GSA.

On 15 August 1990, the GOSIP standard became a compulsory and binding specification for the U.S. government in all of its telecommunication network and computer system solicitations. Implementation of FIPS 146 will help provide a large market for commercial OSI products which will benefit not only U.S. government customers, but commercial customers as well. NIST published GOSIP Version 2.0 in late 1990. The higher versions will be published approximately 18 months apart. The major specifications are listed in table 1.

To assist the transition to GOSIP, U.S. government agencies have begun developing policies and implementation plans dealing with GOSIP procurement and operational issues. Agencies are finding out early that they must deal with registration issues. In the U.S. GOSIP implementation, NIST has delegated operational registration responsibility to GSA. GSA has begun registering administrative authority identifiers (AAI), organization names. and technical objective identifiers for U.S. GOSIP users.

GSA GOSIP Registration

Besides the benefits of protocol standardization that result from GOSIP implementation, an important fallout is that registration of addresses for all federal government end users (administrations and organizations) and names of technical objects will be standardized globally. As a consequence, if anyone needs to locate or transmit information to an agency's location, the uniqueness of the agency's organization and user will be assured. Thus, worldwide communications between users both in and out of the government can be realized. The need for registration was anticipated in the first edition of the *GOSIP Users Guide* [6]. GSA has begun registering Network Service Access Point (NSAP) AAI and has provided a revised chapter on registration for the next version of the U.S. *GOSIP Users Guide*. The following sections of this paper describe the registration procedures as currently being implemented by GSA. They address the question: What needs to be registered?

NSAP Administrative Authority Identifiers

Globally unique NSAP addresses are important in the federal environment, and elsewhere, so that every end system, from personal computer to mainframe, can potentially communicate with each other unambiguously. NSAPs are the first category of objects that GSA has undertaken to register that will assure unique address assignments. The structure of the GOSIP NSAP address is illustrated in figure 1.

The NSAP address consists of two major parts, the initial domain part (IDP) and the domain-specific part (DSP). The IDP is specified by International Organization for Standardization (ISO) 8348/Addendum 2. GOSIP uses a format defined by IS0 international code designator (IS0 6523) in which there is a binary syntax for the DSP. This is signified by an authority and format identifier (AFI) value of 47. The initial domain identifier (IDI) value of 0005 represents the domain which has been assigned to the U.S.

The format of the DSP is not defined by the standard but must be established by the registration authority (i.e., GSA) for the 0005 domain. GSA has decided, during its initial registration of identifiers, to reserve one of the octets within the NSAP AAI block for future allocation. As a consequence, the initial identifiers being assigned are being incremented by 256 in decimal, or 000 100 in hexadecimal notation. Under the registration authority arrangement, GSA has also been assigning, maintaining, and publicizing unique organization identifiers for federal government organizational units, as requested.

Table I. GOSIP Versions Status

U.S. GOSIP Version 1.0 Characteristics

- File Transfer, Access, and Management (FTAM) II
- Message Handling Systems (MHSs) or (X.400) 1984
- Mandatory support of Transport Class 4 and support of Transport Class 0 for Message Transfer Agents (MTAs) connected to public messaging domains
- Mandatory support of Connectionless Network Service (CLNS) and support of Connection-Oriented Network Service (CONS) in MTAs connected to public messaging domains
- Data Link protocols include: High-Level Data Link Control (HDLC) and Link Access Procedure B (L-APB) in conjunction with X.25; and IS0 8802/2 in conjunction with IS0 8802/3, ISO-8802/4. or IS0 8802/5

U.S. GOSIP Version 2.0 Characteristics

- Virtual Terminal (VT) remote terminal access
- End System-Intermediate System (ES-IS) protocol
- Connectionless Transport Service (CLTS)
- Connection-Oriented Network Service (CONS)
- Office Document Architecture (ODA)
- Integrated Services Digital Network (ISDN)

U.S. GOSIP Version 3.0 and Above Characteristics

- Directory services (X.500)
- Interim Network Management
- IS-IS protocol
- VT additions
- Security
- MHS extensions based on the 1988 CCITT X.400 recommendations
- FTAM extensions
- Fiber Distributed Data Interface (FDDI) protocols
- Electronic Data Interchange (EDI) and Transaction Processing (TP)
- Remote Data Access (RDA)

IDP								
AFI	IDI	DSP						
47	0005	DFI	Admln Author	Reserved	Routing Domain	Area	End System	NSel
1	2	1	3	2	2	2	6	1

Octets

Key:
AFL: Authority and Format Identifier
DFI. DSP Format Identlfier
DSP: Domaln Specific Part
IDI Initial Domain Identifier
IDP. Initial Domain Part
NSAP Network Service Access Point

Figure 1. U.S. Government NSAP Address Structure
NSAP is an identifier that uniquely distinquishes one end system from another in a network of systems. The NSAP addressing structure incorporates various numbering schemes and fields to deal with the diverse users of data communications.

Technical Objects

Currently, the technical objects under considerations include: File Transfer, Access, and Management (FTAM) document application profiles, and other technical objects.

The current GOSIP procedures address common objects that must be registered or are likely to require registration. Other objects may need to be registered now or in the future. Among these other categories of objects are: abstract syntaxes, managed objects, application entities and processes, and relative distinguished names. GSA is considering each of these other objects within its registration plans for the future.

X.400 Organization Names

The unique organization names are assigned by GSA for use in constructing originator/recipient (O/R) names used in X.400-based systems. Procedures for requesting the assignment of an organization identifier from GSA have been defined within the latest version of the U.S. ***GOSIP Users Guide.***

It is assumed that all government organizations will be assigned to some ADMDs; thus, the uniqueness of the government's organizational name could be generally assured to be satisfied simply by its relationship to the ADMD.

Originators and recipients of Message Handling System (MHS) traffic are currently identified by a set of attributes which include country name, administration management domain (ADMD), private management domain (PRMD), organization name, organizational unit, and personal name. Message transfer agents (MTAs), which can discriminate down to the level of a personal name, are considered class 3 MTAs. These names are required under U.S. GOSIP X.400 implementation [4].

It is assumed that all government organizations will be assigned to some ADMDs; thus, the uniqueness of the government's organizational name could be generally assured to be satisfied simply by its relationship to the ADMD.

However, there may be circumstances where the ADMD attribute is not used. For instance, a PRMD may be connected directly to another PRMD under a bilateral agreement. In this case, we must be concerned only that the PRMD name is unique within a narrowly defined context (the PRMD-to-PRMD relationship), and wide latitude can be given in the assignment and use of organization names. It would also be desirable for a PRMD attached to multiple ADMDs to enter a single space as the ADMD name. In this case, the PRMD name must be unique within a much broader context. No final resolution has been reached on how to assure appropriately unique PRMD names under U.S. GOSIP. It is expected that a constructed syntax using the identifier,

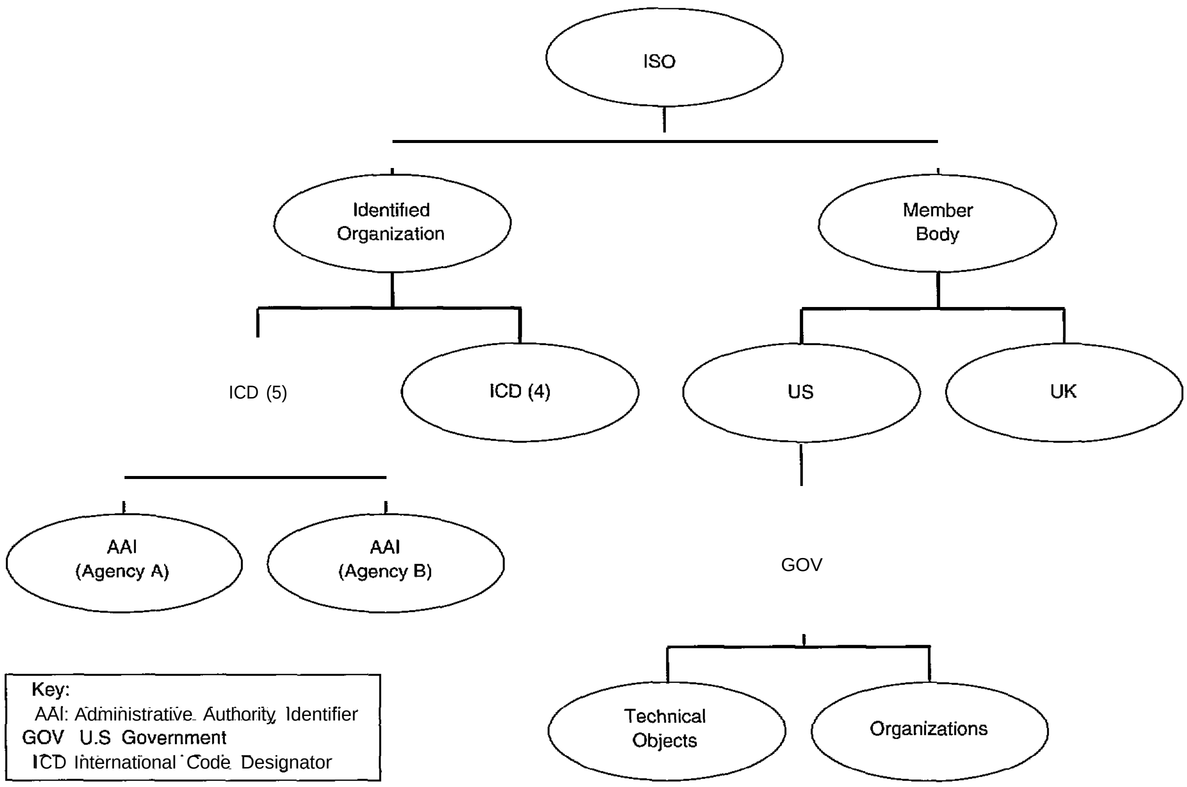


Figure 2. Sample Registration Structure
There are two major naming trees related to GOSIP registration. One is under U.S. country code with value of 840. Another is under International Code Designator with value of 5.

GOV along with the assigned organization name (e.g., GSA.GOV, or GOV+GSA, etc.), will be agreed upon as a generally satisfactory solution.

Registration Trees

There are two major registration trees under GSA's authority (figure 2). One is under IS0 3166, Codes for the Representation of Names of Countries. Under that, the United States has been assigned an alpha-2 code of the U.S. and a numeric code. The American National Standards Institute (ANSI) has assigned the federal government the alpha code of GOV and the numeric code of 10 1. The U.S. government could be unambiguously identified in an international directory as:

{ iso(1) member body(2) US(840) GOV(101)}

Another tree that falls under IS0 is IS0 6523, Structure for the Identification of Organizations. The British Standards Institute has been delegated the authority by IS0 to issue international code designators (ICDs). NIST has received ICDs of 0004,0005, and 0014 and is authorized to issue organization names under those codes. The three codes are used for the Open Systems Interconnection Network (OSINET), government-wide use, and the OIW, respectively. ICD 0004 has been delegated to OSINET. NIST has delegated the authority under 0005 to GSA.

Procedures and Guidance to Agencies

NIST has delegated to GSA the authority to assign values for NSAP AAI under IS0 6523 {iso(1) identified organization(3) ICD(5)}, X.400 organization names under IS0 3 166 { iso(1) member-body (2) US(840) GOV(lOl)}, and Technical Objects under [iso (1) member body (2) US(840) GOV(lO1) technical-objects(O)]. The AAI unique members will start with 256 in decimal or 000100 in hexadecimal. The procedure to assign an NSAP address is described in the U.S. *GOSIP Users Guide.* Unique organization names are provided for use in constructing X.400 O/R names. The procedure for organization name registration is provided in the U.S. *GOSIP Users Guide* following the ANSI registration procedures. In a technical registration, a definition of an object is necessary along with the identification of the object. The numbers for technical object will be assigned in a sub-arc with registration number 0 in decimal under the GOV branch.

In general, the registration process is initiated by each individual agency following the procedure guide and submitting the necessary forms. After GSA's evaluation, if accepted, GSA will send a return letter to the agency head indicating the identifier assigned, effective date of registration, and any other pertinent information. If rejected, GSA will send a letter to the agency head explaining the reason for rejection and requesting alternative assignments. Each agency's registration authority will assign and register its own sub-address space in accordance with the procedures set forth by GSA in the U.S. *GOSIP Users Guide.* GSA will maintain, publicize, and disseminate the assign values of the identifiers unless specifically requested by an agency not to do so.

GSA will maintain, publicize, and disseminate the assign values of the identifiers unless specifically requested by an agency not to do so.

Future Work

In this paper. we have reviewed the U.S. GOSIP versions status and the registration implementation. On 15 August 1990, the GOSIP Version 1 .0 (FIPS 146) became a compulsory and binding specification for the U.S. government in all of its telecommunication network and computer system solicitations. The following versions will be issued about 18 months apart to incorporate new functionality. The implementation of GOSIP will benefit not only U.S. government customers, but commercial customers as well.

In addition to the protocol standardization, an important fallout from the GOSIP implementation is the registration for GOSIP users. The registration procedures established by GSA constitute a minimal set that is considered both necessary and sufficient to serve the U.S. GOSIP users through the first few years of the OSI era. We believe that they are flexible enough to accommodate future needs and have allowed for other objects which may require special registration as standards and procedures develop. Currently unstable issues. such as those involving the apparent uniqueness of organizational names, are not considered serious since fall-back positions can be taken if a problem arises. These issues are what are considered to be a natural part of the progress being made in the implementation of GOSIP.

Currently, there are several registration-related activities under way. For example, in order to resolve national MHS-management domain (MD) issues, a study group was formed within the U.S. Department of State Study Group D. The committee had its first meeting on 17- 18 December 1990. So far, they have established a common base of terminology and semantics, and agreed on the general meaning of the terms "name, address, route, and some technical definitions." They plan to meet every three months. Other efforts including the study of addressing issues related to ISDN and GOSIP networks [7], the NSAP structure study at NIST [4], and the registration handbook and system development [3], are under way. With those efforts by different organizations, we believe the quality of the GOSIP registration implementation will be improved, and serve a key role in overall GOSIP environments.

References

1. Ohle, W., F. Ferrante, V. Hsin, E. Gardner, "U.S. Government Open Systems Interconnection Profile (GOSIP) Status and Registration Implementation," Proceedings *of International Symposium on Message Handling Systems and Application Layer Communication Protocols,* 3-5 October 1990, North Holland Publisher, Amsterdam, The Netherlands.

2. Hsin, V.. E. Gardner, F. Ferrante, 1990, *Registration Procedures Section for the GOSIP Users Guide,* MTR-90W00075, The MITRE Corporation, McLean, VA.

3. Hsin, V., R. Cossa, F. Ferrante, 1990, *Preliminary Standards Implementation Plan,* MITRE WP-9OWOO225, The MITRE Corporation, McLean, VA.

4. *Stable Implementation Agreement for Open Systems Interconnection Protocols,* 1989, Version 3, Edition 1, NIST Spec. Publ. 500-177.

5. *Government Open Systems Interconnection Profile (GOSIP),* 1988, Version 1.0, FIPS 146, NIST.

6. Boland, T., 1989, *GOSIP Users Guide,* NIST Spec. Publ. 500-163.

7. Rudisill, R., 1990, *Addressing Issues Related to Interconnection **of an** ISDN and GOSIP Network.* MITRE W36-L- 183, The MITRE Corporation, McLean, VA.

About the Authors

Victor J. Hsin holds a Ph.D. in Operations Research and Information Systems. He is a member of the technical staff in MITRE's Intelligent Highway Systems Group. Before joining MITRE, he had been with AT&T Bell Laboratories from 1980-1985 and the GTE Corporation from 1985-1988. He is vice-chairman of the Sponsor Open Systems interconnection Support cluster group.

Details about Frank Ferrante are at the end of the first article, "Government Broadband Networking Environment".

A Framework for Selecting FTS2000 Data Services

William J. Kelly

When government agencies transition to FTS2000, they must implement networks that use one or more services from those offered by FTS2000. This article presents a framework for selecting FTS2000 data services based on the type of traffic to be carried by the network. Two types of traffic are highlighted: terminal-to-host and local-area-network-to-local-area-network.

Introduction

FTS2000 consists, in part, of a set of long-distance telecommunications services that government agencies are mandated to use by Federal Information Processing Standard (FIPS) 146. As existing contracts expire, agencies are planning to transition their networks from existing services to FTS2000 services. At the same time, agency requirements for networking are changing quickly, and the design of the networks is also changing to satisfy the changed requirements. In many cases, the decision is not merely to replace existing services with FTS2000 services while leaving the network unchanged; rather, it is to select services that support the evolving network which satisfies new requirements. Because the availability and price of services affect the design of the evolving network, the problem can seem circular. This article begins with first principles-the changing user requirements. From these requirements, we derive types of networks, and then we consider the services that can support these types of networks. For clarity, a particular agency's problems are highlighted, but the principles can be extended to other agencies.

An Agency Example

To make the discussion concrete, we will organize the discussion around an example. This example is occasioned by a particular agency, but is representative of many agencies in the government. The agency in question has a contract with a vendor who is operating a private X.25 Packet Switched Network (PSN) for the agency. The switches for the network are located on agency property and the trunks that interconnect the switches are leased from commercial telecommunications vendors. The network is national in scope and most of the traffic is required to use FTS2000 services when the existing contract expires in a few years. The agency has projected its future telecommunications requirements and expects an increase of several times the current volume by 1995, as well as a change in type of user equipment from mostly dumb terminals today to mostly local area network (LAN)-based workstations and image transfer applications in 1995 and beyond.

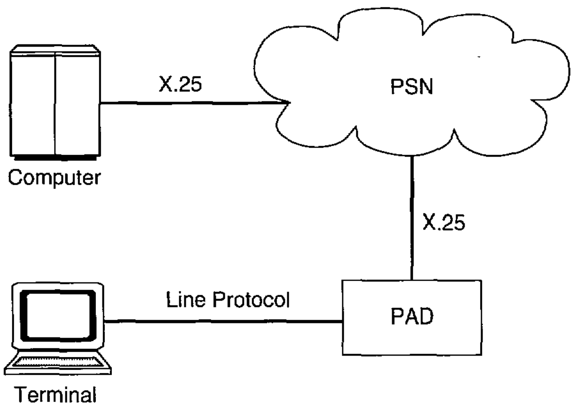
The agency is thus presented with several issues. The existing network must be expanded to satisfy the increasing requirements even though the contract is approaching its end. At the same time, the agency must prepare for the new network that will use FTS2000 services after transition. How should the existing network be expanded? What will the "new" network look like and which FTS2000 services must it employ? The remainder of this article will explore these questions.

User Configuration

The theme of this article is that the nature of the traffic in a network is correlated with the ways that the users interact with the network. This does not say that the physical interface, itself, will change the traffic. Rather, when people do different things with their networked computing resources, both the traffic and the interface may change to accommodate the changed computation practices. We will consider two types of network interface configurations, together with the computation they support and the traffic they generate.

Classic User Configuration

The classic configuration (see figure 1) is all of ten years old, and is, therefore, called *classic* in our fast-changing world. The classic configuration consists of terminals connected to packet assemblers/disassemblers (PADs), which in turn connect to an X.25 PSN, and also to computers connected by an X.25 link to the PSN. The type of computing which occasions the configuration is terminal-to-host communications where the computers do almost all the work, and the terminals are the means to input information and commands, and receive products of computation or database search. In many situations, the terminals are actually implemented on personal computers (PCs) or workstations, but these capable devices are usually reduced to emulating terminals. The type of traffic generated by the computation model and configuration is usually interactive, query-response, or downloading batch jobs. This type of traffic requires reasonably low average delay for single or small groups of packets, but does not usually have very high bit per second (bps) throughput requirements. Because the terminals do not incorporate any mechanism to ensure a minimum error rate, this reliability must be added to the network by implementing error correction (automatically repeated transmission) on the links between PADs and switches. We will associate the PSN with this configuration because the X.25 PSN was developed to convey the corresponding type of traffic.



Key:
PAD: Packet Assembler/Disassembler
PSN: Packet Switched Network
X.25: CCITT Recommendation Covering Packet Switching

Figure 1. Classic Configuration
Terminal with a connection to a mainframe computer through a PSN.

LAN User Configuration

The LAN user configuration (see figure 2) is steadily becoming more prevalent. It consists of computing devices (PC, workstation, minicomputer, or mainframe computer) attached to a LAN. Despite the differences in capability, the devices interact on a peer-to-peer basis. Each device implements a protocol (such as Transmission

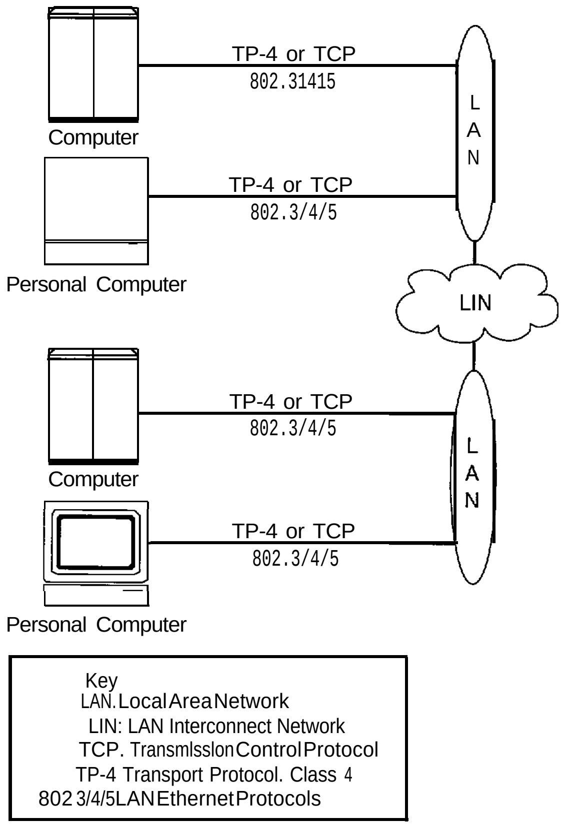


Figure 2. LAN Configuration
PCs and server computers communicating as peers on a LAN and between LANs.

Control Protocol/Internet Protocol [TCP/IP] or TP-4) which ensures the reliable and error-corrected delivery of data to the destination device.

The computation model comprises computing devices exchanging blocks of data for significant processing in each device. The type of traffic generated by this model is usually a mixture of commands to initiate or control the processing and file transfers to exchange the initial and processed data. This type of traffic requires low delay and delivery of large bursts of data. Because the users are typically interacting over a LAN, the wide area network connection should emulate the characteristics of the LAN as closely as possible. Although LANs typically have high data delivery probability, the LAN does not guarantee reliable delivery in the same sense that the error-correcting links in a PSN do. We will call a wide area network, which is optimized for this type of traffic, a LAN Interconnect Network (LIN).

LINs and PSNs

We have distinguished two types of wide area networks, namely LINs and PSNs, on the basis of traffic types generated by particular user configurations. From a design standpoint, what are the differences between LINs and PSNs? PSNs guarantee a certain level of error correction by retransmitting packets on communications links until they pass a correctness test. As line conditions degrade, this means that retransmissions and, consequently, delays tend to increase. On the other hand, processors on a LIN maintain an end-to-end protocol connection which guarantees error correction by retransmission. The performance of such protocols depends on the variability of the expected delay because the protocol does not retransmit until the protocol is “reasonably” sure that the packet is erroneous or lost. This implies a retransmission timeout to detect lost packets, and this timeout must be set longer than the expected maximum round trip delay for the acknowledgment of receipt. If the delay is very variable, then the timeout must be set long, which will degrade performance when the timeout is needed. Thus, a PSN needs a predictable level of errors while a LIN (and LAN) needs a predictable delay. We can highlight the contrast by an idealized comparison.

- Both PSN and LIN need high average probability of delivering data
- An ideal PSN gives a constant minimum probability of delivering data with delivery delay that varies with network conditions

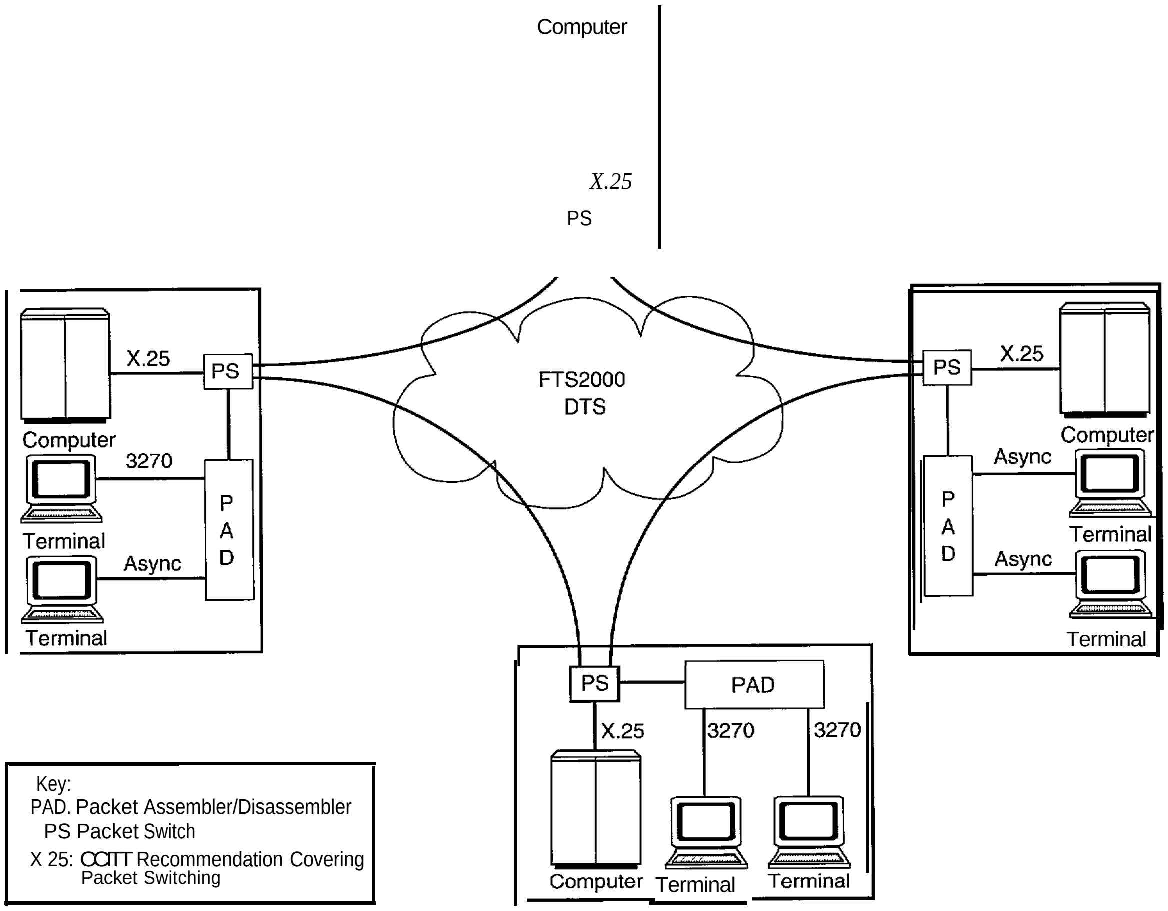


Figure 3. DTS Trunks with Packet Switches
Classic configuration with private packet switching equipment and dedicated transmission service.

- An ideal LIN gives a constant maximum delivery delay with a probability of delivery that varies with network conditions

Available FTS2000 Services

FTS2000 offers a variety of services to support data. Many agencies. including the example agency, have significant requirements for full-time connection to the network. Therefore, for the purposes of this article, we will consider non-circuit switched services only. This reduces the number of alternatives to two.

Dedicated Transmission Service (DTS) offers point-to-point dedicated connections at 2.4, 4.8, 9.6, and 56 kilobits per second (Kbps) and 1.544 megabits per second (Mbps). It is equivalent to the leased lines that the example agency is currently using to interconnect packet switches in the PSN.

Packet Switched Service (PSS) offers X.25 packet switched services with dedicated connections (switched virtual connections and permanent virtual connections). It is equivalent to the service provided by the proprietary packet switches that the example agency is currently using.

Alternate Ways to Transition

For simplicity, we will initially consider two idealized situations and then consider exceptions. In the first case, all users have a classic interface configuration and generate PSN-type traffic. In the second case, all users are attached to LANs and generate LAN-type traffic on a wide area network connecting the LANs. In the first situation, the agency must choose FTS2000 services to implement a PSN; in the second, the agency must implement a LIN. Of course, in the real world, the users may not be so homogeneous. This case will be considered later in this article. Let us consider me PSN decision first.

PSN Implementation

This is the implementation that the example agency is currently using. One alternative is to keep the existing packet switches and use FTS2000 DTS for the interswitch trunks instead of the commercial leased lines. The other alternative is to transfer all the traffic from the private network to the FTS2000 PSS. It would be possible to combine the two approaches for a third alternative, but we will consider only the two. The first alternative is shown in figure 3, and looks the same as the current implementation, except that the vendor of the interswitch trunks (and possibly of the switches) is different. The second alternative is shown in figure 4,

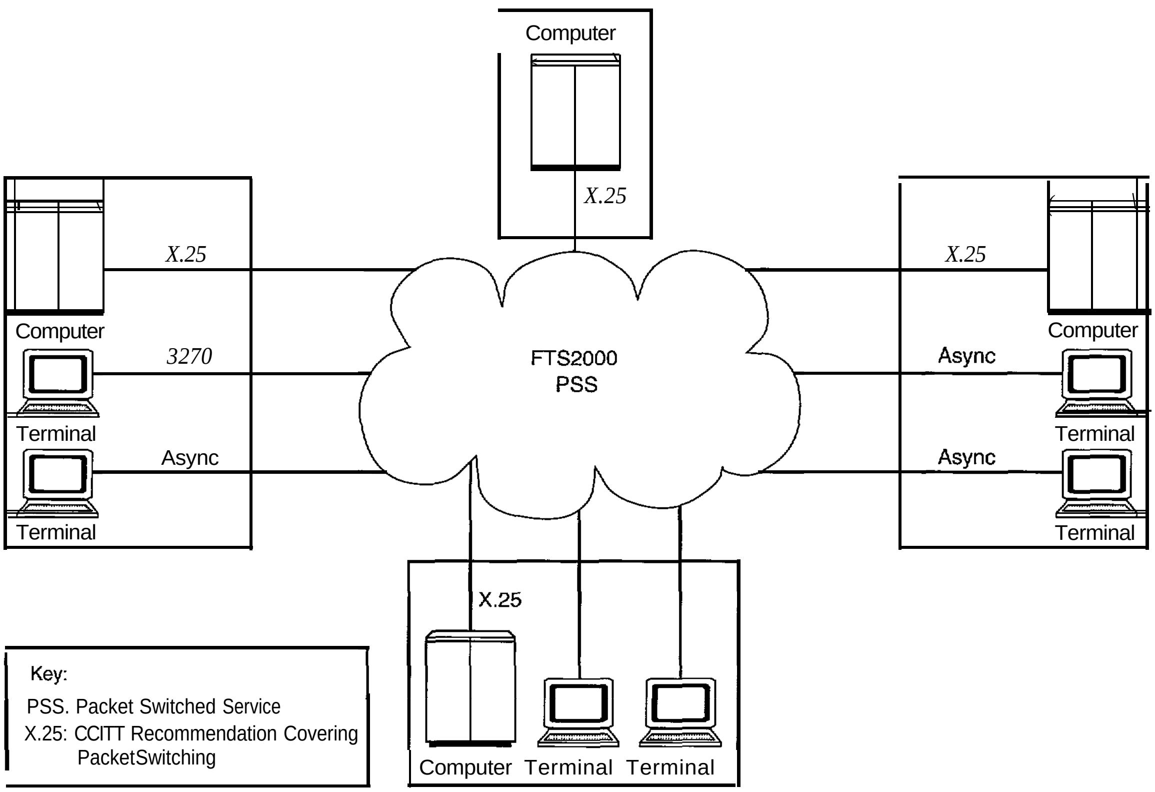


Figure 4. FTS2000 PSS
Classic configuration with packet switching service charged on usage.

and has all terminals connected to FTS2000 PSS PADs and all hosts connected to the PSS via dedicated X.25 lines.

The first alternative has the characteristics and specifications of the private network. This means that the characteristics (reliability, performance, etc.) can be tailored to the requirements of the agency up to the limits of available technology. On the other hand, this freedom is bought at the price of management headaches. The agency must be prepared to manage the network (or the vendor of the network) to ensure that the capabilities of the technology translate to real advantages from the user point of view. The second alternative can offer only the specifications of the FTS2000 PSS, which are good but not tailorable in general to an individual agency. Alternatively, most management effort is expended by GSA and the FTS2000 vendor, so that the agency can devote its management attention to its mission.

The major cost component in a PSN over the life of the network is typically the cost of the communication lines. For both public and private networks, the major determinant of overall cost is efficient use of the communication lines. A network with variable traffic is usually designed for the average traffic to get high utilization of the lines, and then excess line capacity is added for good performance during peaks or bursts of traffic. If the chance of a large peak is low, the final design will not have much excess capacity. If the chance of a large peak is high or unpredictable, the network must have considerable excess capacity for satisfactory performance during such peaks. This excess capacity will increase the cost of carrying the average traffic.

Note that there are two kinds of perdictability-short-term and long-term. Short-term predictability refers to daily or weekly variation. Long-term predictability governs expansion of the network. Capacity normally cannot be added quickly to a network, so the network designer must rely on predictions of when a new application will go on-line (for example), and how much traffic the new application will generate. If these predictions are uncertain, then the network designer must add excess capacity to ensure performance. The designer of a private network must accommodate all these uncertainties. The designer of a public network can rely to some extent on the averaging effect of a large number of users and thus increase the predictability of traffic. If the agency expects its daily traffic variation and long-term growth to be fairly predictable, the designer can probably design the private network to be more efficient than a public one. On the other hand, if the traffic is very unpredictable to the agency, then a public network may be able to use its lines more efficiently than a private network. and it is possible that the public network will offer lower life-cycle costs.

In summary, the choice between a private network with DTS trunks and the public network with PSS is not easy, but the evaluation factors tend to be management-oriented: Do I want to do it myself or have someone else do it? Can I predict my requirements over time, or will I let someone else absorb the risk (at a price, of course)?

LIN Implementation

How can a LIN be implemented? One way is to use a PSN. This implementation is shown schematically in figure 5. The devices called *gateways* translate between the LAN protocols and the X.25 protocols (for example, by providing a PAD). This implementation has two drawbacks. The first is that the PSN delay characteristics are inappropriate for inter-LAN traffic (as mentioned above). The second is that the X.25 gateway processing is complex and can form a bottleneck in itself. The preferred method for interconnecting

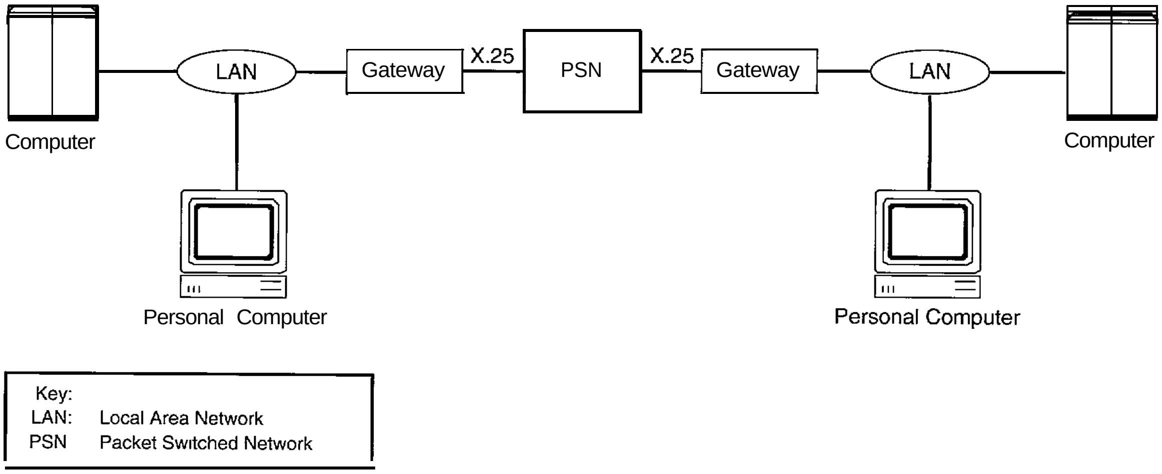


Figure 5. PSN Used as LIN
LAN configuration with L.ANs interconnected using an X.25 virtual circuit.

LANs are devices, such as bridges or routers, that are interconnected with point-to-point links. Routers direct traffic on the basis of addresses from International Standards Organization (ISO) Open Systems Interconnection (OSI) Layer 3 protocols, in contrast to bridges, which use Layer 2 addresses. For large networks with many interconnected LANs, routers are preferred. We assume that federal agencies will typically have many LANs, so routers are assumed in this discussion.

Figure 6 shows a set of LANs interconnected by routers. For good performance with typical LAN traffic. the inter-router links tend to be relatively high capacity (64 Kbps minimum). Therefore, we have shown a T 1 -Multiplexer network which multiplexes the inter-router links over FTS2000 DTS T1 trunks. This is the standard design and the multiplexers allocate fixed bandwidth on the trunks to each link. An implementation with newer equipment is shown in figure 7. A frame relay interface accepts Layer 2 data frames such that different frames have different addresses. A frame relay network conveys these frames on virtual circuits between interfaces. Here the T1 multiplexers have a frame relay interface to which the routers attach and the network of multiplexers becomes a frame relay network. The frame relay interface allows the routers to use virtual circuits for the links instead of dedicated bandwidth. In this case, the traffic on the links is still multiplexed on the trunks, but the virtual circuits allow each link to use the full bandwidth in the absence of contention. This will give superior performance for bursty traffic and allow the LIN to approximate the ideal (LAN-like) characteristics.

Both alternatives use DTS because it is the appropriate service for the point-to-point links. Several wide area carriers have announced the availability of frame relay service. It is very likely that such a service will be available on FTS2000 and would be a very attractive choice, but with the present range of services, DTS is the best choice. The choice between fixed-bandwidth allocation on DTS and frame relay allocation on DTS is primarily a question of technology. The frame relay interface to a DTS network is inherently more appropriate for LAN-to-LAN traffic, but the technology is new and untested, and there is little experience in designing and

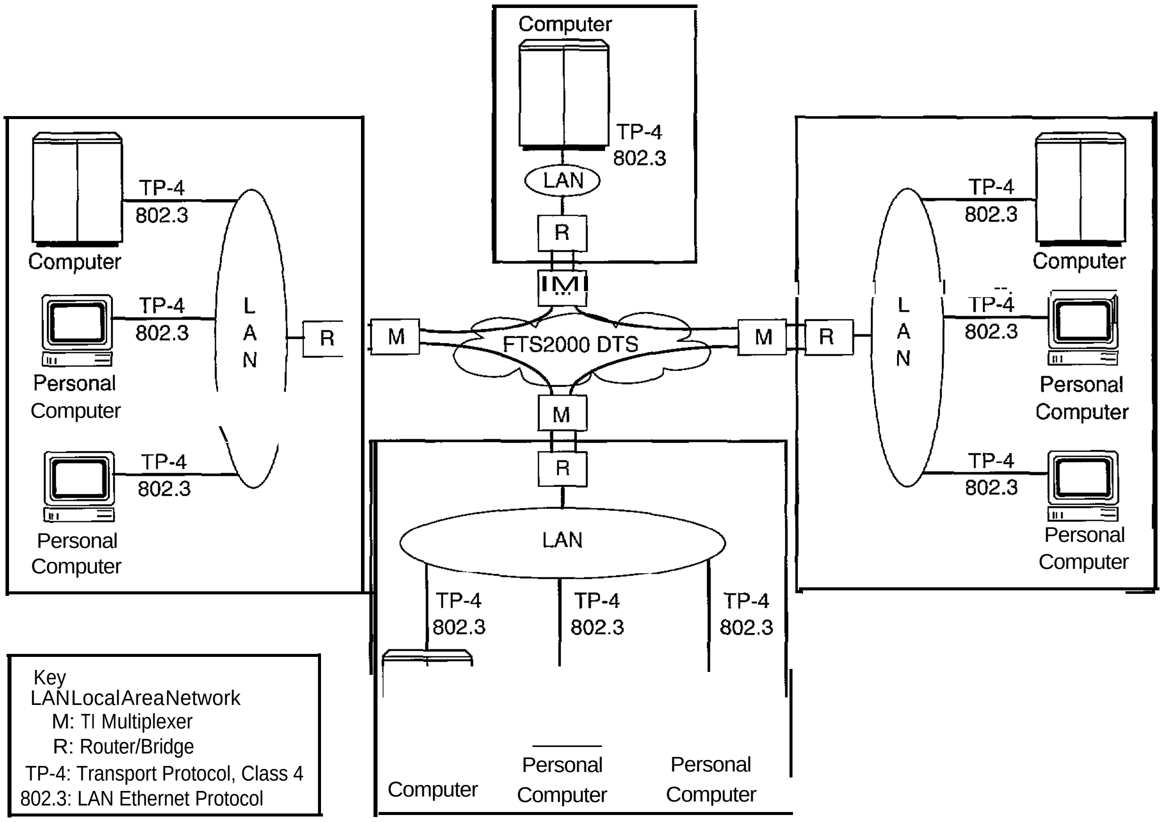


Figure 6. Inter-LAN Routers
LAN configuration using inter-router links multiplexed with fixed bandwidth allocution on high-speed trunks.

managing such networks. There is more experience in designing and using router inter-LAN networks with fixed-bandwidth links, although the art is still not as well understood as the design of PSNs.

Summary of Choices for Homogeneous Networks

Figure 8 diagrams the choices. If the primary user interface is the classic interface, then either the PSS or DTS are possible alternatives with management factors very important to the choice. If the primary user interface is the LAN interface, two LIN alternatives are available, both based on DTS, with technology factors playing an important part in the choice. The decision that a LIN is needed for inter-LAN traffic implies a private network solution with DTS trunks. The agency currently does not have the option of avoiding the management of a private network because FTS2000 does not offer a suitable service.

Choices for Mixed Networks

The preceding analysis is based on an either/or situation. What happens if the users are not all connected in one configuration or the other?

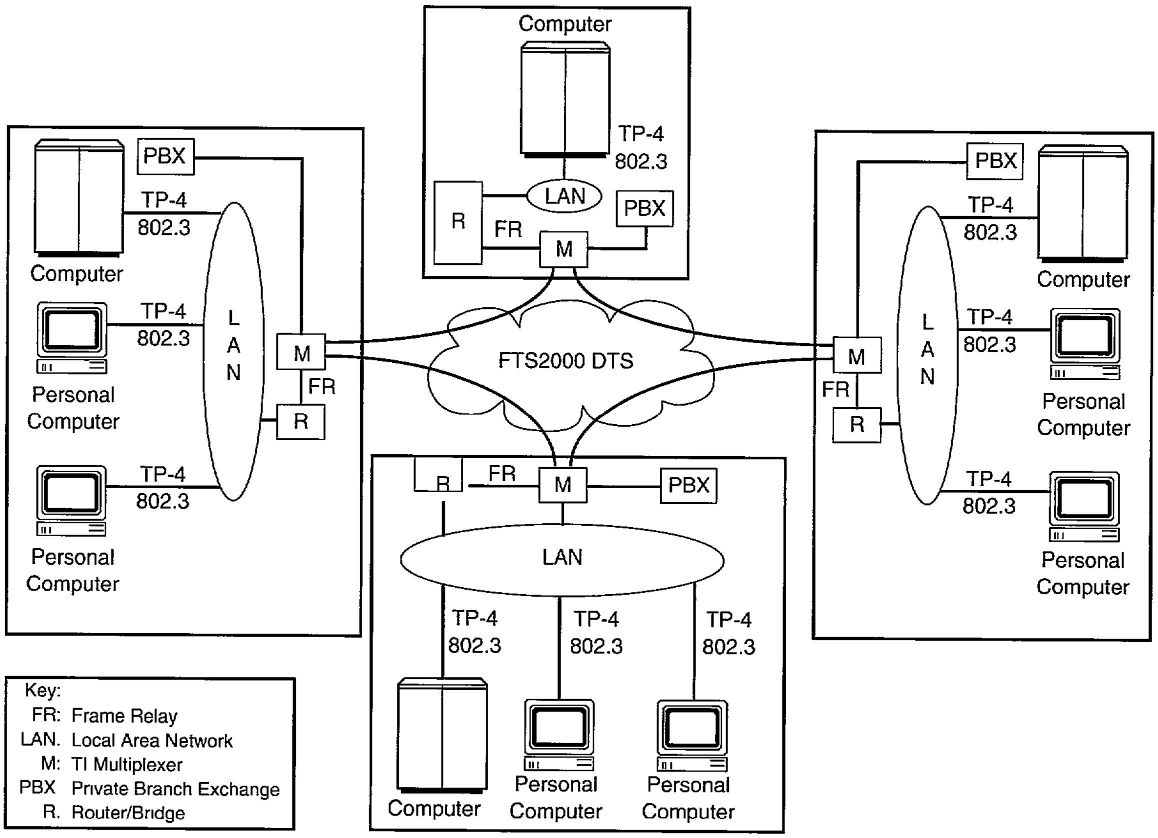


Figure 7. Frame Relay Interface to Multiplexer Network
LAN configuration using inter-router links multiplexed with frame relay virtual circuits on high-speed trunks.

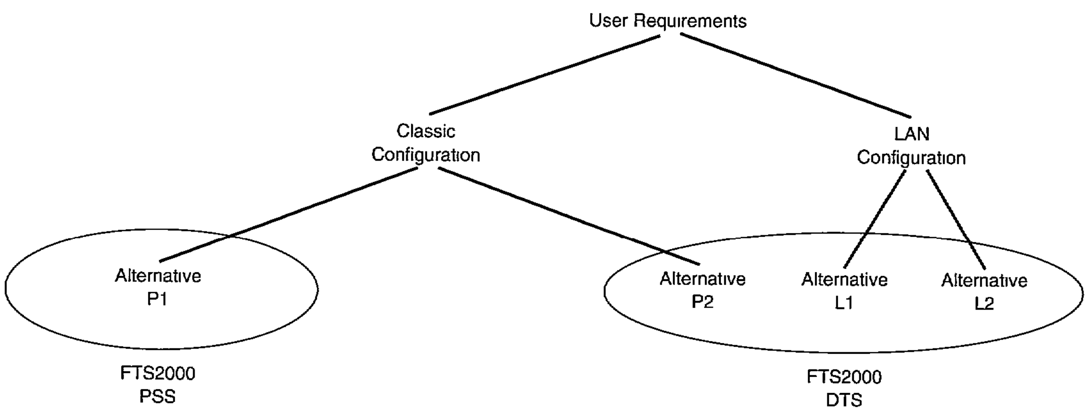


Figure 8. Significant Factors for Choice
Diagram of choices showing relation of alternatives to FTS2000 services.

The simplest answer is that two parallel networks could be used, a PSN and a LIN. Each network would provide the appropriate characteristics for the attached users, and the choices for networks would be as described above. If both networks are private, the trunks from both networks could share Tl DTS trunks. Parallel networks imply that users with one configuration seldom communicate with users of another. In general, we expect this to be true. Users on a PAD expect to talk to applications on an X.25 host, and the applications will be written for that situation.

Suppose the "terminal" users migrate from direct connection to a PAD. to direct attachment to a LAN. There are two possibilities: either the host migrates to a LAN at the same time, or the host keeps the X.25 access. If the host keeps the X.25 access, then a PAD is still needed, and the devices on the LAN must use a PAD (e.g., in a server gateway) and have a requirement for PSN service. If the host migrates to a LAN connection, then a LIN is appropriate. This emphasizes an important point. The network is evolving, as the application programs and consequent traffic load evolve, and choices should be based on a time sequence of events. This topic will be addressed in a future article.

About the Author

Bill Kelly is a lead engineer in the information Systems Engineering Department of the Advanced information Systems Division. His work has been concentrated in analysis. design, and procurement support to projects for civilian agencies of the federal government. He is currently working on the transition of an agency's data traffic from a private network to the FTS2000. Bill obtained his B.S. in Electrical Engineering from Villanova University and his M.S. and Ph.D. in Electrical Engineering from Notre Dame.

Migration to OSI-Based Network Management: Issues for Government Networks

Kris Krishnan

Various organizations within the federal government are in the process of modernizing their telecommunications networks through programs such as the FTS2000 and the Navy Base Information Transfer System (BITS). The thrust of these programs is to integrate voice, data, video, image, and message communications by exploiting the evolving technologies, such as the Integrated Services Digital Network (ISDN), and make the telecommunications systems cost effective. It is becoming increasingly evident that Integrated Network Management is crucial to achieve this objective. This article discusses the complexity and challenge of Network Management (NM) in the evolving telecommunications network environment and points to the need for migration to the International Standards Organization (ISO) Open Systems Interconnection (OSI)-based NM. It also highlights some issues in this context.

Introduction

Network Management (NM) has evolved from a primitive network fault detection and correction function to a sophisticated set of functions covering network administration, operation, maintenance, planning, and evolution of telecommunications networks. Historically, telephone companies considered NM mostly from their perspective as service providers, not from the perspective of the customers. NM was initially confined to detection of network-affecting problems, such as network equipment failure and traffic overloads, and activation of real-time controls such as rerouting traffic or blocking traffic from entering a congested network. During the 1960s and 1970s with the introduction of Stored Program Control Switches (SPCSs) and computerized Operating Systems (OSs), more sophisticated network monitoring, data collection, and network control capabilities became feasible. Since a single OS can monitor and control a large number of switches, OSs make centralized NM possible. As the use of computers became widespread through the early 1970s, remote data communications access was supported through the use of public switched or private leased lines. These networks were small and simple, and NM techniques were primitive with a major emphasis on correcting faults and other problems. Customer diagnostic techniques isolated faults in one of four primary network components: mainframe computers, user terminals, customer-owned data communications equipment, and network. The advent of the modem era of customer-based Network Management Systems (NMSs) began in the mid-1970s when modem vendors began to introduce intelligent test modules that communicated with a main controller located at a central site. This equipment allowed communications managers to isolate failed components better.

Management systems have been built into many types of multiplexers, X.25 packet assembler/disassembler (PAD) equipment, local area networks (LANs) and private branch exchange (PBX) managers, and other network components by the early 1980s. Many of these units also support performance and other functional areas in addition to fault management. Recently, software systems have also been introduced by many vendors and service providers to assist in such areas as configuration management, which includes management of databases of network elements and accounting management. These new systems allow communication charges to be allocated properly to individual divisions within an organization. NM also assists with planning, which requires evaluating the data obtained from the various management activities and reflecting the results to activities such as workforce planning and facility planning.

Existing NMSs are based on proprietary technologies and have interoperability problems. They are not flexible enough to accommodate the needs of evolving integrated telecommunications environments.

There are several trends that point to the need for integrating disparate network management systems through a standardized interface....

Integrated Network Management

There are several trends that point to the need for integrating disparate NMSs through a standardized interface, foremost among which are:

- **Increasing Size of Private Networks:** The private networks of large organizations, such as government agencies and Fortune 500 companies, are very large and continue to expand, placing increasing importance on reliability and quality of service. A failure of some key Network Elements (NEs) can affect large segments or perhaps the entire networks. The key to avoiding this problem is to enable efficient communication and interpretability between distributed network management facilities.

- **Increasing Complexity of Private Networks:** In addition to the fact that NEs, such as the PBXs, Packet Switches (PSs), and LANs, are becoming more varied and sophisticated, there is a strong multivendor trend whereby various types of equipment supplied by a variety of vendors exist within a single network. This situation makes it difficult for NM personnel to cope with new technologies if a separate management system is used for each equipment type or vendor. The shortage of skilled maintenance technicians is also becoming a serious problem. In addition, network maintenance becomes expensive due to the need for stocking spare parts, inventory control, etc., for varied NMSs.

- **International Networks:** In many instances, private networks have expanded into international networks and NM has become a worldwide management of a network. Because of the variations in NMSs, particularly the communications protocols associated with these systems from one country to another, management of private networks spanning over many countries can become very complicated and inefficient.

- **Multiple Carriers and Hybrid Networks:** Traditionally, the products, services, and tools for managing the carrier facilities were simply unavailable to customers. Hence, these were some limitations that the customers had to be aware of and work around but did not have the ability to change, influence, or utilize. This situation is changing rapidly in the post-divestiture era since the competition between the carriers has started to increase, and customers want to exercise more control over their networks. Private networks need a flexible interface to switch between different carriers for leased, long-haul services. Besides, customers want to exercise control over the leased facilities for such functions as network reconfiguration or adapting to changing traffic requirements. This is feasible through new concepts such as AT&T's Software Defined Network (SDN). These features require interaction between the NM facilities of carriers and private networks, including private network access to carriers' databases.

- **Integrated Services:** The distinction between telecommunications and information processing is becoming blurred due to new technologies such as the Integrated Services Digital Network (ISDN), which integrates voice, data, and video communications services. Common Channel Signaling (CCS) is being introduced to allow faster and more complex signaling between different parts of the network leading to more sophisticated intelligent network services; Switched Multimegabit Data Service (SMDS) and Fiber Data Distributed Interface (FDDI) provide high-speed connectivity between LANs which are geographically separated. Broadband ISDN (BISDN) will enable communications at much higher speeds, from 100 megabits per second (Mbps) to 600 Mbps and provide higher speed data, video, and image communications and has potential for many new and sophisticated applications in the integrated environment such as videophone and high-resolution facsimile. These new technologies will have a significant impact on NM. Some of these are already being implemented while others are expected to proliferate by late 1990s. Integrated Information Transfer Network (IITN) is becoming a vital resource to large and small organizations. A failure of even a small segment of this network can result in significant loss of productivity and/or revenue. Therefore, today's communications managers have to take a pro-active role in managing this resource.

OSI-Based Network Management

It is evident from the preceding discussions that existing NMSs, based on proprietary technologies, are inadequate to meet the needs of evolving IITNs. As we move into the 21 st century, we need a robust NM infrastructure to integrate and manage the corporate communications resources in an efficient manner. The key to this challenge is to plan and develop an integrated NMS which is flexible and adaptable, as shown in figure 1. Because of the largely distributed nature of the enterprise networks with multivendor products and multimedia services,

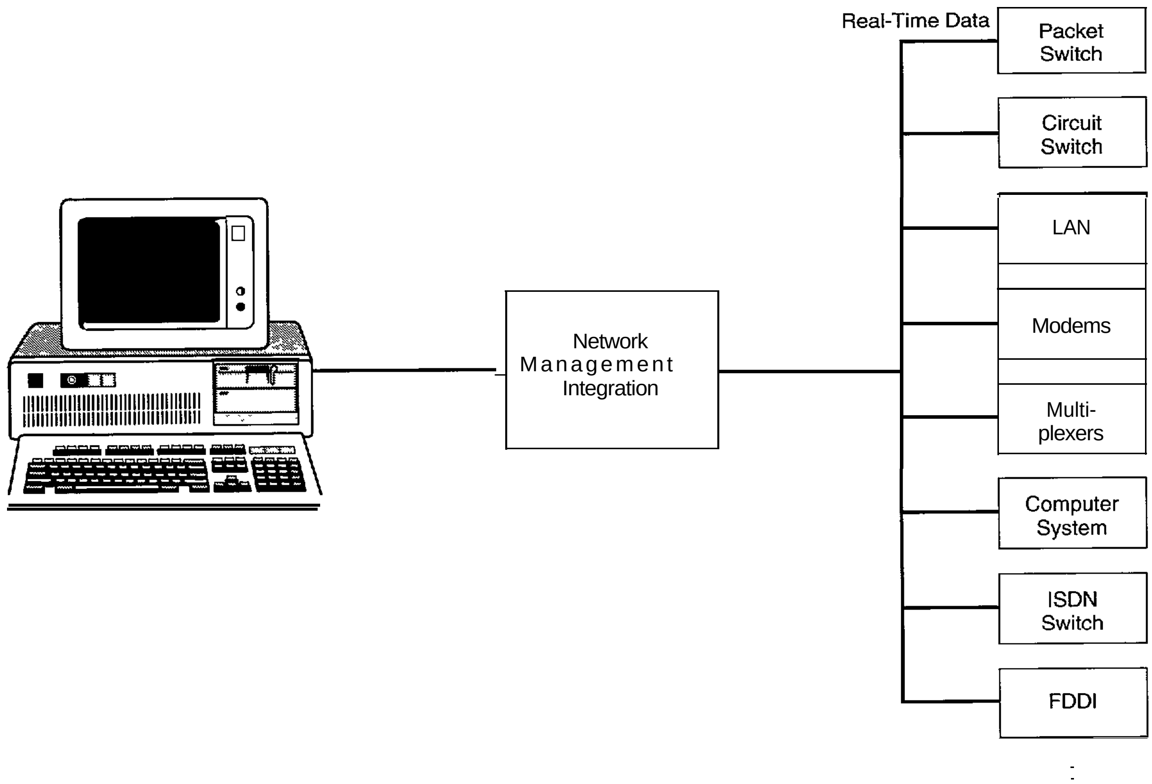


Figure 1. Network Management Integration
Figure I is a conceptual illustration of an IN management disparate/distributed component of an Enterprise Information Network.

we need a standardized approach to this network management challenge which is precisely what IS0 OSI-based NM strives to accomplish.

The aim of OSI NM standards is to allow inter-operability and true integration between a large number of separate, isolated components in the IITN.

The aim of OSI NM standards is to allow interoperability and true integration between a large number of separate, isolated NM components in the IITN. The concept behind this approach is illustrated through a simplified model shown in figure 2. NM is accomplished through a set of interactions between one or more managing processes and one or more managed processes. A managed process is responsible for one or more managed objects. A managed object is an abstract representation of any resource from a management perspective. It may represent a physical resource (e.g., a multiplexer) or it may represent something that is itself an abstraction (e.g., a customer account).

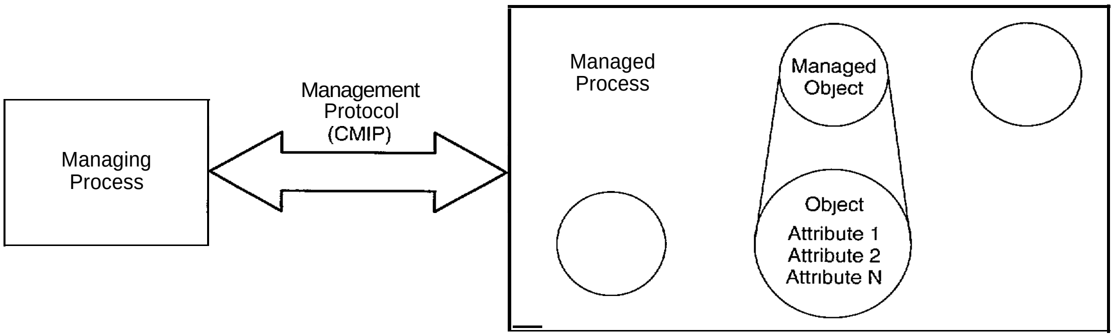


Figure 2. OSI-Based Management Concept
The OSI Management concept draws heavily on the principles of object modeling. Standardized management protocols and service are the keys to Integrate management.

The shared conceptual repository of all managed object information is known as the Management Information Base (MIB). A managing process manipulates information by an agent process, using the services provided by the Common Management Information Service (CMIS) and conveyed in the Common Management Information Protocol (CMIP) operation, to perform NM functions. These functions include fault management, configuration management, accounting management, performance management, and security management.

Within the MIB, information is largely embodied in terms of the attributes of the associated objects. Each attribute of an object has a value. The CMIS provides services to retrieve (Get), modify (Set) attribute values, and create (Create) and delete (Delete) objects. In addition to objects, information may be embodied in notifications using the CMIS Event Report. Notification signals the occurrence of some event to an object, e.g., a static change. These four services provide the capability for manipulating and reporting managed objects. The Action service is used to cause specific physical activities, e.g., to request the execution of a diagnostic test.

The OSI management standards defined above are expected to be completed by 1992. Many parts will reach final standards status before that time. Standards defining CMIS and the CMIP protocols have already achieved this status.

Implementation Agreements

As the standards incorporate numerous options to suit the needs of many nations and industries, the standards do not directly lead to the development of interoperable products and so OSI management standards are just the first step to NM interpretability. The next step is to specify implementation agreements. Implementation agreements serve several needed purposes. There are many protocol options at each OSI layer. Implementation Profiles (IPs) specify which particular options must be implemented, include conformance statements to these selections. and may discuss testing concerns related to conformance. IPs also specify many implementation details not included in the standards.

An NM Special Interest Group (NMSIG) has been working since July 1987 within the OSI Implementors Workshop (OIW) and is writing implementation agreements for the emerging OSI management standards. The European Workshop for OSI Standardization (EWOS) and the Asian Workshop for OSI Standardization (AWOS) are two other implementation workshops. All three have agreed to resolve any differences and submit their input to ISO, which will then issue international standardized profiles (ISPs).

Because there are several options built into the OSI NM standards, it is imperative to define a common subset of these standards for use between various management entities within a large distributed system.

Conformance Testing

Another important step in helping to support interoperability is conformance testing. This determines whether an implementation complies with the relevant OSI standards and profiles, and thus greatly increases the probability that different implementations can interoperate, although it cannot ensure this. No conformance testing for OSI management exists yet.

Migration Issues

There are several issues that need to be addressed in the context of migration of government communication networks to OSI-based NM. These include:

- Definition of an evolution concept
- Development of a common subset of OSI NM standards
- Development of an acquisition strategy

An evolution concept for transitioning from the current environment to an integrated environment, tailored to meet the needs of individual organizations, will enable a smooth and orderly process for migration. This concept should take into account the NM requirements of the current and future networks as they evolve toward a target of integrated OSI-based management systems. Issues, such as centralized versus distributed architectures for NM facilities and security, also need to be addressed in this regard. Since the current disparate NM systems, such as Transmission Control Protocol/Internet Protocol (TCP/IP) SNMP-based systems and proprietary voice NMSs, will continue to exist for a long period, provision must be made for interpretability between OSI and non-OSI-based systems. Future systems will include FDDI and ISDN and integration of these is a major issue. Once a concept has been solidified, a migration plan can be developed taking into account the developments within the standards bodies and commercial sector.

Because there are several options built into the OSI NM standards, it is imperative to define a common subset of these standards for interoperability between various management entities within a large distributed system. This includes CMIS/CMIP, and definition of managed objects. Definition and development of MIB is another major issue. There are initiatives by some government organizations already under way. For example, the Department of Defense (DOD) has formed a NM working group under the Data Communications Protocol Standards (DCPS) Technical Management Panel (DTMP) for this purpose. MITRE plays an active role in helping DOD achieve this goal of NM standardization.

In parallel with these efforts, it is also necessary to monitor, and if necessary, influence various standards activities pertaining to NM such as is in the OIW and NMSIG. Even though there is some government participation through MITRE and government representatives, increased participation will be necessary in the future especially to address government-unique requirements such as security. Influencing the standards bodies can help avoid costly developmental components by enabling the government to procure commercial off-the-shelf (COTS) components for NM.

OSI NM products based on IS0 standards are not expected to be widely available until the mid-90s. Commercial products available today, such as Digital Equipment Corporation Enterprise Management Architecture (DEC EMA), IBM NetView, and AT&T Unified Network Management (UNMA). offer only proprietary solutions. However, due to the rapid pace of evolution of some networks, NM plans and implementations cannot be delayed until IS0 OSI-based NM products are available. One approach to this issue is to acquire products that have the potential and vendor assurance for migration to OSI standards when it is feasible. In this context, the product should have the flexibility to integrate new components based on new technologies such as ISDN, PBXs, and FDDI. Some products, such as the AT&T Accumaster product family, offer this type of solution.

Summary and Conclusions

As the government networks evolve to an integrated voice, data, and video communications environment, NM will play a crucial role in achieving the basic objectives of the modernization process. The future network environment will be highly distributed and very complex due to numerous technologies such as ISDN and FDDI. Also, deregulation of the telecommunications industry has resulted in a highly competitive environment and has necessitated increased customer control through interaction between the carrier and customer NM facilities. These and many other factors point to the need for migrating to IS0 OSI-based NM, a standardized network management approach for interpretability between diverse systems, flexibility to change in accordance with user needs, and adaptability to evolving technologies. A systematic approach is needed for the government networks to migrate from the current disparate NM systems to an integrated NM based on the OSI NM framework. Several issues need to be addressed in this regard. Increased government participation in standards bodies is also required to influence the NM standards ultimately to enable the government organizations to acquire COTS NM components.

About the Author

Iyengar (Kris) Krishnan is a member of the technical staff in MITRE's National Telecommunications and Data Systems Department. Prior to joining MITRE in 1987, Kris spent eight years with AT&T Bell Laboratories, Alcatel, and ITT, primarily working in the areas of telecommunications network design, planning, and architecture. He holds a B.E. in Electronics and Communications Engineering, an M.E. in Computer Science, and a Ph.D. in Electrical Engineering.

Intelligent Network Evolution Status

Vladimir Nikanorov
Therese Metcalf

This article reviews the principles and ongoing development of Intelligent Network (IN) concepts, standards, and implementations. Intelligent Network (IN) is a new technology that offers a phtform for implementation of communications services and provides the basis for general interoperability between information and telecommunications systems. The core of the paper is dedicated to a high-level description of the evolving CCITT IN modeling concept. The article also reviews the issues arising from IN implementations in the United States, within the European communities, and the federal telecommunications environment.

Introduction

"Intelligent Network" is a label adopted by standards organizations to describe a concept that permits users of telecommunications services to specify and "design" their own customized services. In general, IN is not synonymous with "intelligence" in a network; rather, it describes a specific use of the intelligence associated with the latest advances in telecommunications, such as distributed data processing, database management. expert systems, new circuit and packet switching techniques, broadband digital transmission and switching, frame relay, and common channel signaling [1].

IN as a term and its basic concept were conceived at Bellcore initially, as Bellcore's answer to demands for flexibility and cost-effectiveness in providing new services such as calling card, 800, and virtual private network. The conventional implementation of these services requires a huge expense in changing or discarding the stored-program controlled switching machines. These difficulties are avoided by an innovative approach based on the use of central databases, specialized computer support, and common channel signaling. The concept has become generalized and is now known as the Intelligent Network.

The telecommunications industry, service providers, and equipment manufacturers have turned to the IN concept as a result of prior experience with Integrated Services Digital Network (ISDN). ISDN and access arrangements provide users with integrated data-voice terminals and access arrangements, but fall short in supporting conveniences, such as 800, telephone number portability, calling card, and closed user group, equally with the voice and data bearer services. The experience revealed that integration of service and freedom of access, theoretically available from ISDN, are, in fact, limited by the rigidity of ISDN services, complexity of the ISDN-supporting switches, and costs.

Unlike ISDN and other upcoming technologies, such as Broadband ISDN (BISDN), frame relay, and Switched Multimegabit Data Service (SMDS) that must replace current switching equipment, IN is intended to complement existing telecommunications schemes [2]. In effect, IN protects existing investments and increases the effectiveness of future investments in telecommunications over public networks. IN also offers major advantages to end users. With IN, users would be able to make and receive data calls as conveniently as voice calls anywhere in the world using their unique telephone number. Service features, such as personalized announcements, customized billing, and voice recognition, as well as a variety of user custom-tailored features, could be available in an IN environment.

The majority of IN developments have taken place within telecommunications companies (common carriers, equipment manufacturers, etc.) in the United States and in Europe. The results of this work are being submitted to standard bodies. The current U.S. contributions are being approved in the American National Standards Institute (ANSI) Tl S 1.1 -IN subworking group; the European developments are being discussed and approved by the European Telecommunications Standards Institute (ETSI). In turn, the work from these two groups is submitted to the International Telegraph and Telephone Consultative Committee (CCITT) Study Groups XI-4 and XVIII for possible adoption as international recommendations.

Capability Set One (CS- 1) [3] is the first set of IN specifications to be approved by T 1 S 1 -ETSI this year, and will be approved by CCITT in 1992. CS-1 defines the basic IN architecture and accommodates the relevant known telecommunications protocols with some elements of the open systems architecture. CS-1 also standardizes a limited number of the core service features for introduction under IN such as virtual private network, 800 service, and Universal Personal Telecommunications. Subsequent sets of IN specifications, CS-2, CS-3, ... CS-n, will support advances in information processing merged with advances in telecommunications. These future IN specifications will incorporate the full extent of OSI principles, open distributed processing, application portability, artificial intelligence, and object-oriented technology. The evolving CSs will enable users to define customized services based on IN-provided access to known technologies, such as ISDN, BISDN, frame relay, and SMDS, and will support the access to new technologies that may become available in the future. The CCITT recommendations and corresponding domestic standards will also provide detailed guidance to implementation of IN.

Intelligent Network Objectives

IN is a concept that offers a variety of tools to meet specific objectives: ease of creation and manipulation of services including management services, and the ability to interoperate and interconnect a variety of networks from which the services are available. These tools are offered to end users or service subscribers, network and service providers, and equipment manufacturers.

From the end-user perspective, IN is based on two major principles. The first is subscriber control over standard service features. This enables a subscriber to define a variety of custom services. The second is the independence of IN services from network implementation. IN services may span several telecommunications networks.

From the perspective of both service providers and network providers, IN is a mechanism that allows the creation of new user services through the use of standard building blocks stored in centralized databases and supported by dedicated systems. This process is independent from the network mandatory functions such as switching, signaling, and protocols invocation. As a result, the network service implementation is no longer directly driven by services. Networks will evolve for the purpose of perfecting their mandatory functions without affecting the assortment or quality of existing and new service combinations.

Equipment manufacturers view IN as providing separations among four aspects of operation: applications, logical, functional, and physical. These separations make a network independent of what brand of physical elements (software or hardware) are used. As a result, IN physical elements could be procured from different vendors.

Intelligent Network Conceptual Model

The IN conceptual model [4] shown in figure 1 is based on the philosophy that supports a high degree of interoperability and flexibility in the introduction and implementation of services. To achieve this objective, the application, the logical, functional, and physical features of telecommunications, currently available from monolithically designed switches, have to be

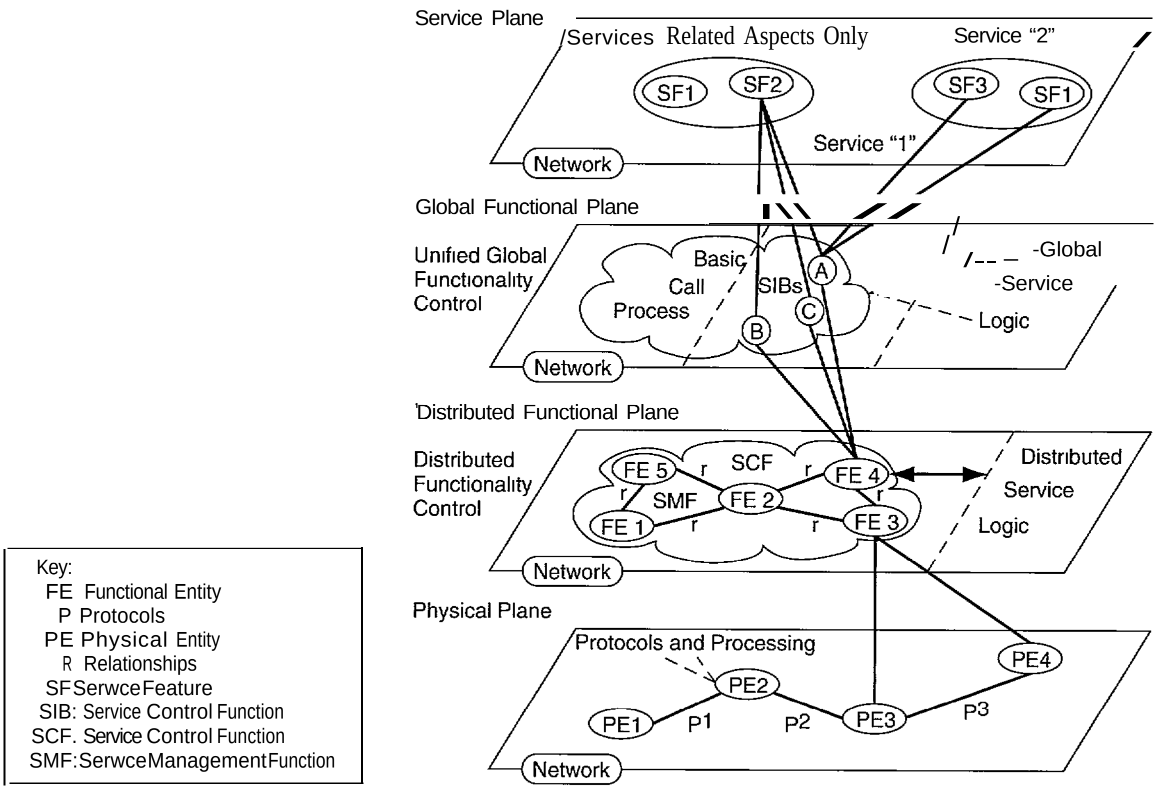


Figure 1. IN Conceptual Model

This figure shows a modeling concept for the Intelligent Network. Each plane represents a different model (or a view) of a telecommunications network: the service plane views it from applications' perspectives: the global and distributed planes from the functional perspective; and the physical plane from the hardware/software distribution perspective.

divorced, distributed, and treated independently of each other. To be consistent with this philosophy, the IN concept is based on a four-plane model. The model includes the service plane, the global functional plane, the distributed functional plane. and the physical plane. Figure 1 illustrates the CCITT IN conceptual model. Each of the planes contains a logical view of the entire IN, These four views, however, represent different levels of abstraction and are aimed at different stakeholders. The common element presented in each of the planes is the service logic that uses functionality in its own plane to provide functions in the next higher or lower plane.

The service plane [5] represents an exclusively service-oriented view of the entire IN. This view is deliberately narrow and contains no information whatsoever about the implementation of the services in a network. The service plane contains generic services features that could be assembled in a set according to a user's requirements and describes a specific service for this user. The service plane is indifferent to the nature of a service whether it is a telecommunications service or one that serves management. Therefore, all the services, including the management services, are seen by the end user on this plane.

A number of service features have been developed for the upcoming standard specification, CS- 1; more are being considered as the understanding of IN capabilities becomes more sophisticated. Examples include an array of familiar and unfamiliar features, such as call allocation, calling number identification, user prompter, access authorization, shared trunking, remote access, customized billing, and Public Switched Telecommunications Network (PSTN) overflow. (A PSTN overflow feature would allow a user of a private network to transfer calls onto the public network if the trunking facilities of the user's private network are unavailable.)

The dilemma faced by the IN service plane developers is how to identify the service features. One of the obvious ways to analyze the existing services is by gradually decomposing them into: service features. service independent building blocks (SIBs), functional entities, and eventually into functional entity actions that would operate physical entities. This approach is known as the top-down method. The danger of this approach is that an end user may be limited ultimately in the assortment of available services because the features used for the services assembly would be based on previously defined services.

An alternate approach would be to develop the services features using the bottom-up method. This method considers available technologies (i.e., BISDN, frame relay, SMDS) as the basis for the actions in the physical plane. Functional entities, SIBs, and finally a set of generic service features are gradually derived from these actions. The danger of this approach is that the end user may be limited by the technological capabilities available at the time.

Both methods are being considered by IN developers. An answer may be in a combination of the two approaches, which would provide a process of identifying service features independently from existing services and technology.

The global functional plane illustrated in figure 1 is limited to a view of IN as a single monolithic network, a sort of virtual machine programmable by customers from the services plane [6]. The global functional plane is populated with the basic units of IN modularity, the SIBs. They play the role of elementary services in order to assemble the service features in the service plane. The customers "see" monolithic SIBs through standard interfaces; they are not concerned with. and cannot see, the actual SIBs' functions in a network. SIBs developed thus far for CS- 1 include elementary services components such as announcement, billing/charging, service data management, clock, call gap, queuing, authorization, and translation.

For example, the translation SIB should be used in all of the service features that require address modifications. Features like the PSTN overflow service must include the translation SIB for modification of the address information from a private numbering plan to a destination address in the public networks. The PSTN overflow may also require inclusion of the authorization SIB to verify that the calling party is authorized to use the public network. The billing/charging SIB would be needed to facilitate the charges associated with the public network, and so on.

The distributed functional plane illustrated in figure 1 provides another level of abstraction for the IN model based on distribution of functions throughout the entire IN. The distributed functional plane contains functional entities. The distribution of functional entities in a network represents a static model of IN. The functional entities' interactions in an instance of a service execution would provide a dynamic model of IN.
The functional entities are a result of the SIBs' decomposition. They are described in the distributed functional plane and implemented in the physical plane where each functional entity represents a group of functions allocated to a physical entity. The functional entities serve the physical plane in two capacities: they provide the service execution and the service creation management. The functional entities identified for the upcoming standard specification CS-1 are shown in the table 1 below.

The previous paragraphs describe how the user manipulates the service features and how the service features are decomposed to the service independent blocks, and these into the functional entities. Each functional entity, in turn, provides a set of functions. The resulting structure is shown in figure 2. Functional entity actions are defined in a series of subset models such as the Call Connection Model (CCM), Basic Call Model (BCM), and Basic Call State Model (BCSM). The major actions that take place within the CCM are the actions between the Service Control Function and the Service Switching Function. The Service Control Function uses its own group of functions, called the services logic, to control other functional entities in each instance of a service. The Service Control Function logic "opens" and "closes" "sockets" or "windows" for "conversations", and exchanges information with other functional entities. For example, if a network performs the 800 service and one of the tasks is to provide the number translation, the service trigger must open a socket after the service switching function detects the dialed digits. Then, the Service Control Function and the Service Switching Function will engage in a "dialogue" for the duration of the Service Switching Function request of the routing information and the Service Control Function response.

Table I. Functional Entities (FEs) Identified for the IN CS-Z

SERVICE EXECUTION FEs	SERVICE CREATION MANAGEMENT FEs
Call Control Function (SCF)	Service Creation Environment Function (SCEF)
Service Switching Function (SSF)	Service Management Agent Function (SMAF)
Service Data Function (SDF)	Service Management Function (SMF)
Specialized Resource Function (SRF)	

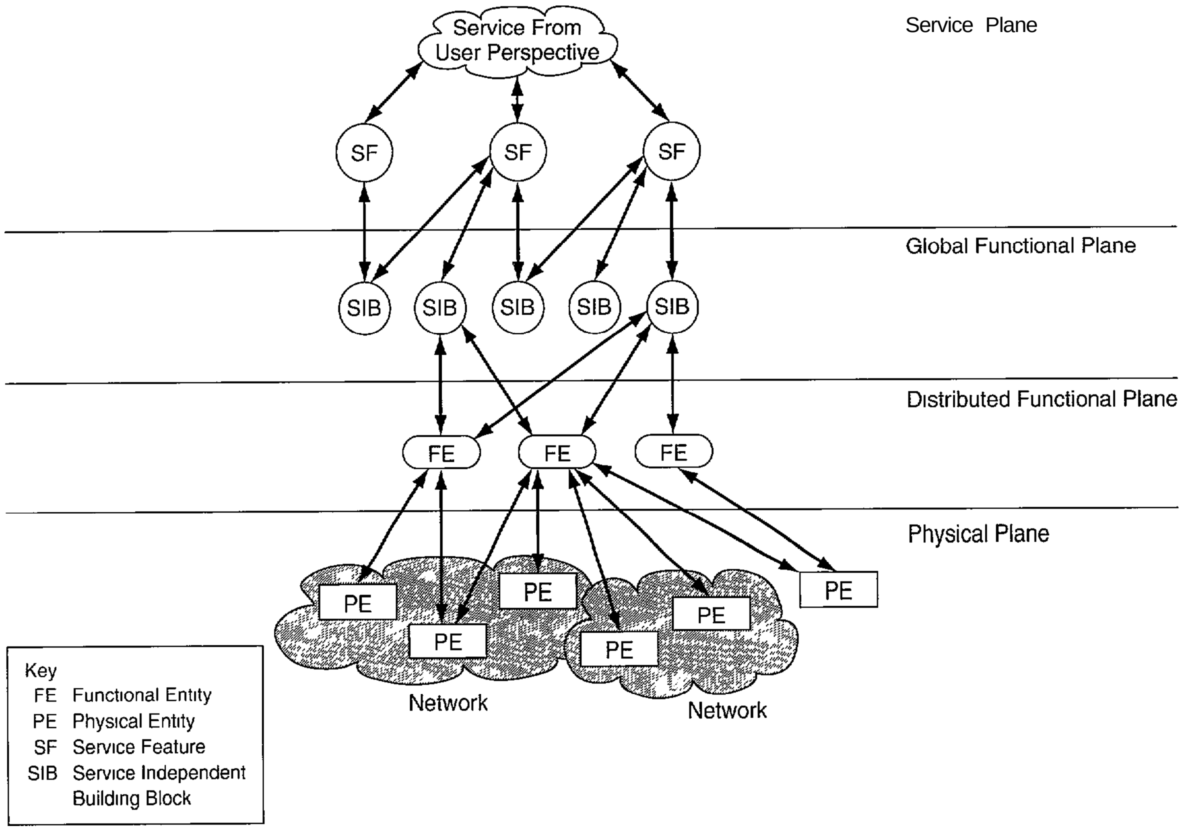


Figure 2. Functional Decomposition of Services
Each of the four independent models (planes) of IN uses the functionality on its own plane to provide functions in the next plane. The service user chooses service features (SFs) for a desired service. These selected SFs are decomposed to service independent building blocks (SIBs) on the global functional plane and to functional entities (FEs) by the distributed functional plane. Specific physical entities (PEs) on the physical plane (switches, databases, networks) recognize the FEs and organize them into actions that constitute the service chosen by the user.

If a service call requires the invocation of basic call procedures in a network. the Service Control Function functional entity works with the BCM. a part of which is the BSCM. The BSCM recognizes call instances, identified in the model as points in call (PICs), and provides the conditions for a chosen path. Examples of these conditions are idle, busy, status waiting, alert, conversation, and release. The Service Switching Function is responsible for a connection setup. It opens a socket at a certain PIC when a trigger event is received by the trigger check point. Examples of the trigger events are release, off hook, setup, alert, answer, busy, and time-out. They must correspond to a specific PIC to satisfy the trigger check point. The combination of a PIC and a trigger event prompts the Service Switching Function to open one socket or another and to control the actions appropriate for an instance of a service. In reality, the Service Switching Function action is more involved. The Service Switching Function communicates with various functional entities, determining connection legs, sending and receiving service primitives, and interacting with the management functional entities. While the distributed functional plane serves the physical plane by providing the functions distribution among physical entities, the physical plane maintains strict independence from the distributed functional plane and the other three planes of IN.

The physical plane has yet another, deliberately narrow and constrained view of IN. This plane recognizes single physical entities, such as switches or databases, and conglomerates of physical entities, networks such as ISDN, packet switched, or Personal Communications Network (PCN). The physical plane model concentrates on the telecommunications equipment capabilities and addresses features such as switching, signaling distribution and signaling protocols, and database contents and capacities. The purpose served by these features is of no interest to the physical plane. This limitation is applied equally to the IN view of single physical entities and to its perception of networks. Despite this strict independence. the physical plane is under an obligation to provide physical entities for certain functional entities and to make available the use of the network protocols to operate the functional entities. The physical plane, in general, describes which of the functional entities are implemented in each of the physical entities. Since the IN physical model is separated from the model for services, actual processes that would take place on the physical plane and in the physical network implementation are not impacted by the applications that either were changed or implemented to satisfy the service subscribers.

Implementation Aspects of IN

While the ink is still drying on IN conceptual documents, major telecommunications companies are announcing IN implementation plans. IN-type concepts could not have come at a better time. With the cost of physical transport falling steeply and demand shifting from ISDN-type services towards IN-based services, such as 800, virtual private networks, and universal personal telecommunications, many see IN offerings boosting revenues with relatively small incremental investments.

Two diverging implementation approaches come from Bellcore and from AT&T. The Bellcore implementation, called Advanced IN (AIN), offers an IN implementation restricted to the regional telecommunications and to voice services. AT&T aims at long-distance traffic and advanced voice and non-voice services. Bell Atlantic Corporation, Ameritech. US West, and Nynex have already scheduled trial AIN offerings for this year. which include the 800 service, virtual private network, and customized billing. They plan to tariff the IN services early next year by taking advantage of Bellcore's pre-standard IN releases [7]. The IN implementation efforts are supported by a variety of equipment manufacturers, such as AT&T Network Systems, Northern Telecom, DSC, and Fujitsu of America. The equipment manufacturers provide both the IN-related hardware enhancements (service control points and adjuncts) to the network's functional architecture and SS7 Signaling Transfer Points that simplify the use of SS7 for IN purposes.

In Europe, the situation is even more explosive. Despite the fact that the European national operators are already pursuing their own independent plans, the growth potential based on the IN offerings is estimated to reach $5 billion after 1996, which amounts to 2 percent of the market. The Europeans. however, are very concerned about the current IN situation. A study prepared by the European Telecommunications Consultancy Organization [7] points out that Bellcore's AIN model targeting local traffic and focusing on the voice service may not be appropriate for most of Europe where the division between the local and long-distance services is nonexistent. The European Telecommunications Consultancy Organization contrasted Bellcore's AIN with the diverging AT&T IN approach and argues that the AT&T approach may ultimately prove best for Europe. There is no consensus on what the final U.S. implementation scheme will be. IN's separated services, call control, and switching functions free network components and resources; the implication in the U.S.-regulated market environment could be pervasive. For the first time, the IN would straddle the old boundaries among equipment manufacturers, public regional and long-distance networks, and private networks, leading in a new era of competition between diverse market players.

A complex thread of new regulatory and technological benefits and limitations would come into play in the U.S. when the IN implementation becomes widespread. In this environment, the federal government will play a unique double role as the regulatory authority and as a potential user. The federal government operates a large private network composed of disparate systems of varying ages and needing integration. Integration of federal telecommunications may take on a new meaning with the adoption of IN principles. Availability of equal access to the IN high-technology features and services would allow "logical" rather than physical integration of the federal telecommunications.

One of the major advantages of IN is equal access. Equal access is implicit in the IN concept, but under current implementation plans, it is not readily available for private networks. This may prove to be a major issue. The issues of equal access are already producing attempts to rebalance public and private networks roles. Rushed implementation efforts, in the absence of national and international standards, force the proliferation of proprietary solutions, such as AIN, which negatively affects the ability of private networks to take advantage of the IN features. Emerging ANSI, ETSI, and CCITT standards neither anticipated the level of interest in IN usage for private networks nor the readiness of the public market for immediate implementation. Thus far, these standard bodies have failed to address the issue of equal access by users of private service providers.

If the currently implemented IN is indeed aimed only at the public network providers, private network providers should consider it as an opportunity for a broad rethinking of the existing balances and roles of current participants in today's telecommunications. They should focus on development of innovative regulatory approaches, IN applications for private networks, and on interworking standards. These future regulations and standards should deliver the advantages of IN to private providers independently of proprietary solutions.

References

1. CCITT, 1988, CCITT Study Group X1-4, Question 1 Document.

2. Bellcore Special Report, September 1990. SR-NWT-001754, Issue 1, Section 1.4.

3. *IN Capability Set Number One* (CS-l), January 199 1, CCITT Report, Working Document 4-16, Question l/Xl, SWP 1.

4. *The IN Baseline Document*, January 1991, CCITT Working Document.

5. *Sub-Technical Committee NA6* document, DTR/NA-6001, January 1990, ETSI Working Document.

6. *Communications Week*, January 199 1, C. M. Publications, Inc., Manhasset, NY.

7. *Communications Week,* February 199 1, C. M. Publications, Inc., Manhasset, NY.

About the Authors

*Vladimir Nikanorov is a member **of** the technical staff in the Advanced Information Systems Division. He holds a B.S. and an MS. in Electrical Engineering and is currently pursuing his Ph.D. in telecommunications. He has 20 years of experience in the area **of** telecommunications including development **of** SS7 and ISDN protocols, in design and testing networks, and in satellite communications. Vladimir has worked extensively in the standards development arena, participating in both CCITT and TI standards committees. Currently, he is involved with the TISI.1 working group charged with the development of Intelligent Network protocols.*

*Therese Metcalf is a member **of** the technical staff in MITRE's Advanced Information Systems Division. She received her B.S. in Computer Science from the University of Maryland, and recently completed requirements for an MS. in Technical Management at the Johns Hopkins University. She spentjive years in the U.S. Air Force; her last assignment was with the 2164 Communications Squadron, maintaining and utilizing mainframe and small computer systems. Therese came to MITRE in 1988. She is currently assigned to the GSA Telecommunications Planning Tusk, which is planning the Integrated Federal Telecommunications System for the next five years. She is also assigned to the Video Services Working Group, obtaining and evaluating the agencies' video service needs.*

GOSIP Transitioning Issues

Richard J. Nieporent, Ph.D.

The Government Open Systems Interconnection Profile (GOSIP) specifies an inter-operable suite of Open Systems Interconnection (OSI) protocols. Since Federal Information Processing Standard (FIPS) 146 now mandates the use of GOSIP within the federal government for all new procurements and major upgrades of existing computer networks, government agencies are faced with the daunting prospect of having to transition to GOSIP This paper discusses some of the major technical and acquisition-related issues that must be addressed by these agencies when implementing GOSIP.

On 15 August 1990, Federal Information Processing Standard (FIPS) 146 took effect. FIPS 146 mandates the use of the Government Open Systems Interconnection Profile (GOSIP) by government agencies for all new computer network procurements and major upgrades of existing networks. Although this day was heralded as a momentous occasion by the trade press, there was no mad rush on the part of government agencies to be the first to implement GOSIP.

One reason for this reluctance is a lack of understanding on the part of these agencies as to the benefits, costs, and risks associated with transitioning to GOSIP. They feel that there is a whole set of technical, marketplace, and acquisition issues that must be resolved before they will be willing to transition to GOSIP. Thus, they are afraid that far from being a panacea, implementing GOSIP now will cause more problems than it will solve.

Since a major activity of MITE is acquisition support, in order to help our sponsors carry out the GOSIP mandate, it is necessary for us to understand the issues and problems associated with transitioning to GOSIP. This article will identify the major GOSIP transitioning issues, and examine potential solutions to these issues that will help government agencies decide when and how to implement GOSIP in their networking environment.

Is GOSIP Real?

One of the major concerns being expressed by some government agencies is that GOSIP is not real, i.e., it is not possible to implement GOSIP now. They say that the Open Systems Interconnection (OSI) protocols are just a set of (still incomplete) paper standards, and that there are no OSI products available from vendors. Further, they state that it has taken ten years for the Department of Defense (DOD) Transmission Control Protocol/ Internet Protocol (TCP/IP) to become mature and be accepted by the user community, and that it will take at least the same amount of time for OSI protocols to reach maturity. Therefore, it would be irresponsible to specify the use of GOSIP until it has at least reached the current state of maturity of the DOD protocol suite.

While it is not possible to discount totally the maturity argument, in reality, this concern is based on a lack of a current knowledge of the commercial marketplace. Even a cursory look at vendor offerings immediately shows that a large number of OSI products exist today. Every vendor has an X.400 and File Transfer, Access, and Management (FTAM) implementation. In addition, most have network, transport, session, and presentation layer software.

More importantly, the biggest vendors (IBM, DEC, Hewlett Packard [HP], UNISYS, Sun, BULL, etc.) are committed to making OSI work. They are making a major effort to develop and market new OSI products. If one reads the trade papers, it appears that new product announcements are being made almost on a daily basis. Thus, the GOSIP mandate (FIPS 146) does not require the vendors to provide a whole new set of networking products; it just reinforces an existing commitment to provide OSI products.

Therefore, the real issue is not whether OSI products are available, but whether the vendors can provide a complete suite of GOSIP-compliant protocols now. In one sense, it is not possible to do so yet. The necessary conformance tests are just becoming available, so that no vendor can truthfully claim that its software is fully GOSIP compliant. However, this problem is not as critical as it might first appear. FIPS 146 only mandates that after August 1990, all future procurements for communications software must implement GOSIP protocols. For most procurements, the actual implementation will take place a number of years later, and the necessary conformance testing will have been developed by then. In the meantime, government agency networks can be developed using existing OSI products that will be replaced by GOSIP-compliant versions as they become available.

Is GOSIP Needed?

Another major concern of some government agencies is whether OSI/GOSIP is really needed. Critics of OSI say that it is too little, too late, because an open suite of protocols already exists. Just about every computer vendor provides a UNIX operating system (OS) with the DOD TCP/IP protocols implemented in the OS kernel, and the Simple Mail Transfer Protocol (SMTP), File Transfer Protocol (FTP), and Remote logon (TELNET) protocols implemented as applications in user space. In addition. IP routers are available from a large number of vendors for internetworking local area networks (LANs) and wide area networks (WANs).

OSI protocols will provide enough value-added functionality, as well as interoperability, to justify the evolutionary replacement of other network protocols.

These critics of OSI also point out that not only are many more TCP/IP products currently available than OSI products, but, in addition, the number of new TCP/IP products that are being released exceeds the number of OSI products being released. Thus, they argue, why transition to OSI protocols when the DOD protocols already provide an "open" (in the sense of being nonvendor-specific) suite of protocols that are widely available in the marketplace.

In reality, the fact that there are a large number of DOD products currently available is not that relevant because the vendor community has also committed itself to providing complete suites of OSI protocols. Enormous sums of money have already been spent by the vendors to support the OSI standardization process, and to develop OSI products. It is extremely unlikely that the vendors will abandon this effort after all of the money that they have spent, and are continuing to spend, on OSI. The only way that they can recoup these costs is to market reliable and robust OSI products.

A more important issue in deciding whether GOSIP is needed is the functionality provided by the protocol suite. Current OSI applications offer increased functionality over comparable DOD protocols. In addition, a large number of new OSI applications are being developed that will provide functionality not available in any other protocol suite. Thus, OSI protocols will provide enough value-added functionality, as well as interoperability, to justify the evolutionary replacement of other network protocols.

The existence of a large number of TCP/IP products will mean, however, that it will be necessary for vendors to develop strategies for coexistence that will allow users to support both OSI and TCP/IP products (as well as proprietary products) on the same network. Thus, users will not have to replace existing networks with OSI networks. Rather, they will begin to implement OSI products that will interoperate with existing network products. Eventually, they will replace their existing networks with OSI networks, but it will be an evolutionary process. This will minimize the cost of transitioning to OSI.

Will GOSIP Meet All User Needs?

Versions 1 .0 and 2.0 of GOSIP provide the "basic" application communications services of file transfer, electronic mail, and remote logon (virtual terminal). This is the same functionality that is provided by the DOD protocols that are implemented in the Berkeley UNIX 4.3 operating system. Thus, if an agency is currently only using these applications, then GOSIP will provide an equivalent capability. However, if an agency needs additional functionality, such as directory services, security, network management, transaction processing, or distributed database access, then GOSIP will not currently provide the functionality needed. (With the exception of network management, the DOD protocols also do not provide this functionality.)

Since GOSIP is an evolutionary standard, the National Institute of Standards and Technology (NIST) is planning to add additional functionality to future versions of GOSIP. Version 1 .O of GOSIP provides file transfer and electronic mail applications. Version 2.0 of GOSIP adds virtual terminal protocol, office document architecture, end system-to-intermediate-system (ES-IS) routing, and Integrated Services Digital Network (ISDN). This new functionality will be mandated in October 1992. Similarly, additional functionality for security, network management, directory services, enhanced messaging (Message Handling System [MHS]- 1988), transaction processing, remote database access, and IS-IS routing has been identified for Versions 3.0 and 4.0 of GOSIP. These versions of GOSIP will be mandated when the protocols specified have become international standards.

Thus, even if the current version of GOSIP does not meet an agency's needs, future versions of GOSIP should meet those needs. An agency, therefore, at a minimum, should evaluate GOSIP during its normal system upgrade cycle to identify when the functionality it needs will be available in GOSIP. This will enable the agency to plan its transition to GOSIP systematically.

Is There a Performance Impact Due to GOSIP?

Conventional wisdom states that the performance of OSI protocols will not be as good as DOD protocols. The greater number of layers of OSI, as compared to the DOD protocols, would seem to lend credence to the belief that the OSI protocols will provide a lower level of performance than the DOD protocols. Surprisingly. however, performance testing of OSI protocols has shown that there is little, if any, performance degradation. The reason for this appears to be that protocol performance is affected more by how the protocol is implemented than the inherent efficiency nf the protocol. Thus, for example, the Sun implementation of OSI protocols uses the same operating system calls as its TCP/IP implementation, and the performance of the two protocol suites are similar.

Also, if the OSI protocol implementation uses common memory to avoid copying the protocol data units from one protocol layer to another, then it appears that there will be little degradation due to the additional OSI layers. The one major exception to this is in the use of ASN. 1 in the OSI presentation layer. However, it is not necessary to use ASN. 1 to translate between homogeneous environments; therefore, there would be no performance impact for data transfer between identical systems.

The transitioning technique used could have a much greater impact on performance than the protocols themselves. Thus, for instance, the use of an application gateway can introduce a large delay in a file transfer, independent of the efficiency of the OSI protocol. Also, a transport service bridge will impact performance since the protocol data unit must first be sent to the bridge, and a new checksum calculated, before the protocol data unit is sent to the destination host.

What Products Are/Will Be Available to Implement GOSIP?

Users will have available two categories of GOSIP products. The first category will be complete GOSIP stacks that include protocols for all seven layers. For agencies implementing new systems, these products will provide the functionality to support the development of applications based on the GOSIP protocols.

The second category of products will be transition tools. These products will enable a user to add GOSIP functionality to an existing network. Included in these products will be application gateways and transport service bridges. Application gateways will enable hosts using GOSIP protocols to pass information to hosts using another protocol suite. For example, the FTP/FTAM application gateway will enable files to be transferred between hosts that use the DOD and GOSIP protocols. Transport service bridges will enable hosts to run GOSIP application layer protocols on top of networks that use a different transport and network protocol. For example, Request for Comments (RFC) 1006 will enable hosts running GOSIP applications to send data over networks using the TCP/IP protocol suite.

By implementing OSI protocols in the UNIX operating system, Berkeley UNIX 4.4 will provide the environment to enable OSI to become as widely used as TCP/IP protocols.

Currently, all major vendors (IBM, DEC, HP, Sun, UNISYS, BULL, etc.) have OSI products available, and are planning to implement additional products that are GOSIP compliant. In addition, third party vendors, such as Touch and Retix, have complete suites of OSI protocols that can be ported to a given vendor's hardware. Also, Version 6.0 of the IS0 development environment (ISODE) software is available. ISODE contains a full suite of OSI protocols. In addition, it includes RFC 1006 for running GOSIP over TCP/IP, and application gateways to transfer files and electronic mail between DOD and GOSIP systems.

Finally, the next version of Berkeley UNIX (4.4) will contain a full suite of OSI protocols and transitioning tools, in addition to the current TCP/IP protocols. By implementing OSI protocols in the UNIX operating system, Berkeley UNIX 4.4 will provide the environment to enable OSI to become as widely used as TCP/IP protocols.

What Transitioning Techniques Should Be Used?

A number of transitioning techniques exist which go from a non-GOSIP to a GOSIP environment. Each of these techniques has benefits

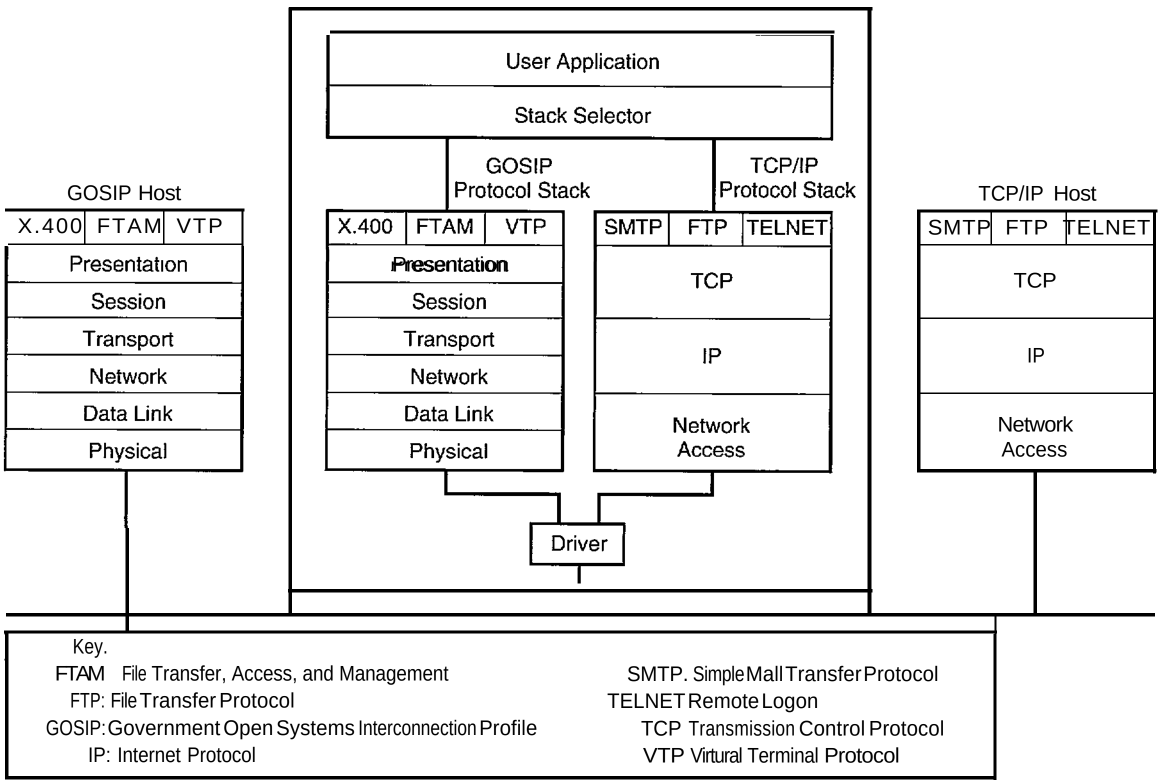


Figure 1. Dual Protocol Host
A dual protocol host allows a user to communicate with any other user that implements either the DOD or the GOSIP protocol suite.

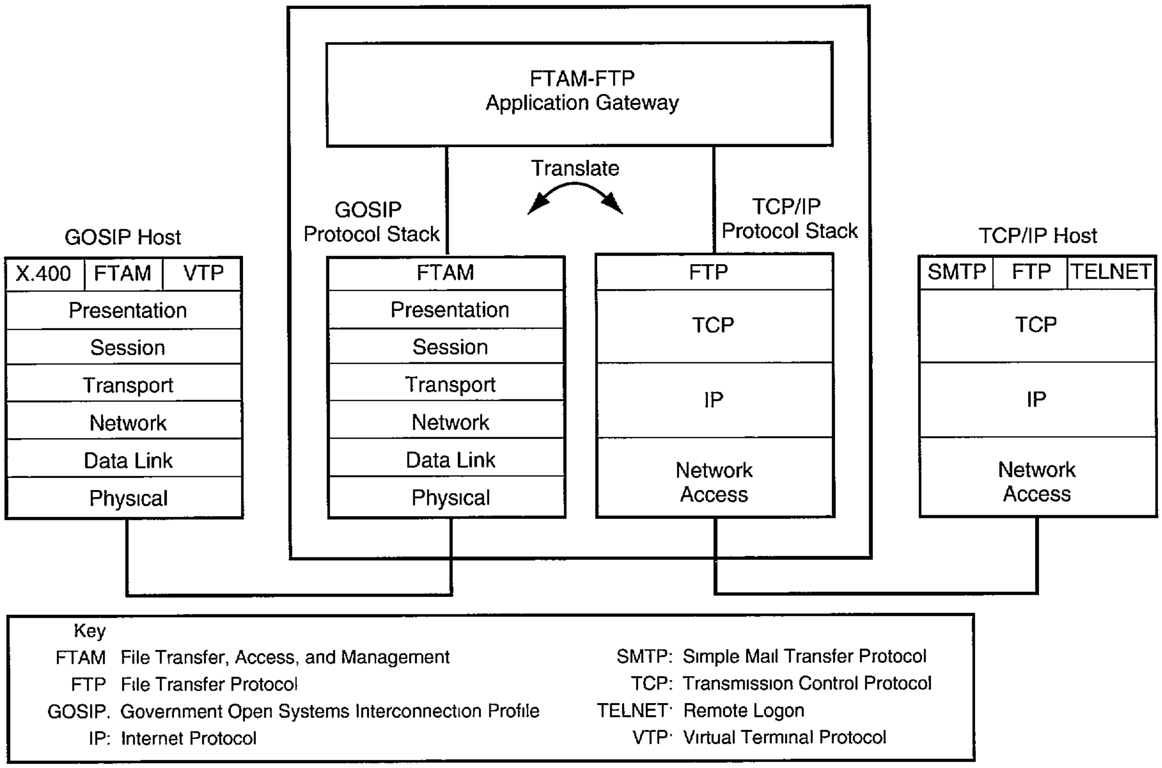


Figure 2. FTAM-FTP Application Gateway
The application gateway translates between two different application protocols. By using the FTP/FTAM application gateway, files sent using the DOD FTP file transfer protocol can be read on a system using the OSI FTAM file transfer protocol and vice versa.

and drawbacks associated with its use. For instance, a dual protocol suite (as shown in figure 1) can be implemented that will enable a new system to communicate with an existing system using the current protocol suite and with another new system using the GOSIP protocol suite. This technique provides the benefit of a transparent user environment; however, the user will be constrained to use only the functionality of the existing protocol suite until all systems transition to GOSIP protocols.

Application gateways can be implemented to allow systems using different protocol suites to communicate with each other (as shown in figure 2). This is also a transparent transitioning technique. However, it is limited to supporting only the common functionality of the applications from each protocol suite, and, in addition, it may not provide the necessary responsiveness for the user's application.

Similarly, such techniques as transport service bridges (figure 3) and network service tunnels (figure 4), can be used to transition to GOSIP. However, these techniques also have limitations. For instance, transport service bridges leave the network configuration with a single point of failure, and cause increased over-

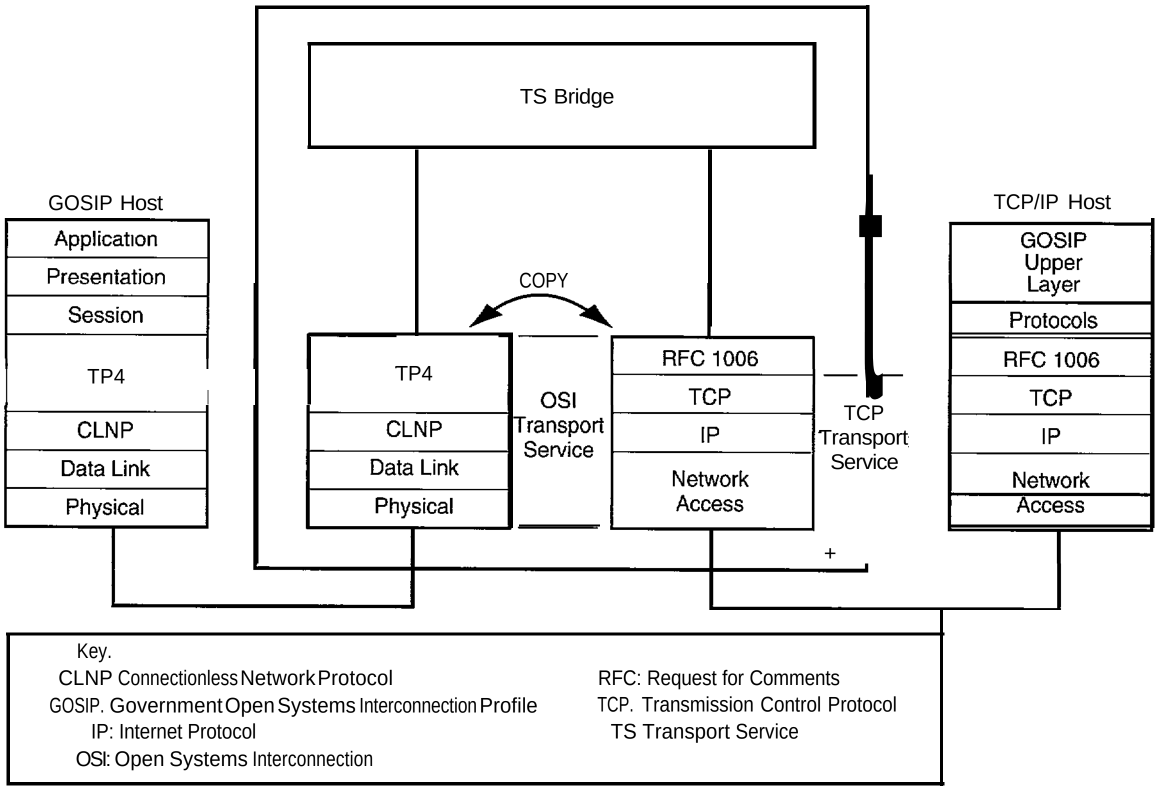


Figure 3. Transport Service Bridge
The transport service bridge allows GOSIP application layer protocols to be run on top of a DOD TCP/IP network. The user on the GOSIP host will be able to communicate with the user on the DOD host using GOSIP application layer protocols.

head due to the requirement to recalculate the checksum. In addition, network service tunnels, while providing a transparent technique for sending GOSIP traffic over non-GOSIP routers, do not provide for interoperability between hosts using different application layer protocols.

Thus, there is no perfect transitioning technique; each one has its benefits and drawbacks. The government agency should, therefore, select the technique that has the least impact on its operation.

What Is the Best Strategy for Implementing GOSIP?

The government agency must be aware of the different strategies available for implementing GOSIP, and choose the strategy that is appropriate for its environment. Thus, for example, the agency must evaluate the advantage of going directly to a full GOSIP implementation versus the use of a mixed protocol implementation. In the former case, only a single transition has to take place.

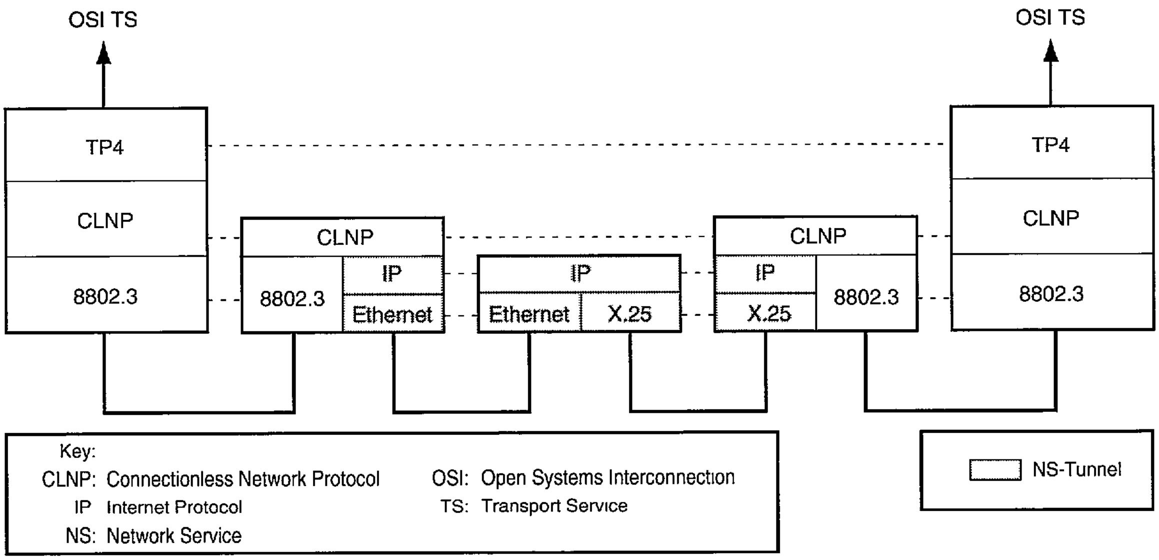


Figure 4. Network Service Tunnel
A network service tunnel allows OSI hosts to use the DOD IP protocol for internetworking. This diagram shows two OSI routers communicating with each other using an IP router as a "tunnel."

In the latter case, there first has to be a transition to the mixed protocol suite, and later, a second transition has to take place to replace the non-GOSIP protocols with GOSIP protocols.

While a single transition is certainly preferable to multiple transitions, it may not be practical. For instance, the OSI internet (IS-IS) routing protocol is only a draft standard at the present time. Thus, it will be necessary to modify the internet routing software in the future, when the IS-IS protocol is specified in GOSIP. In this case, it would make sense to implement first GOSIP applications over a non-GOSIP network, and later transition to GOSIP when the internet routing protocols are available and mature. On the other hand, if the applications the agency needs are not yet available in GOSIP, it may make sense to implement first a network using GOSIP protocols to allow the agency to gain experience with these protocols. Later, the agency can transition its applications to GOSIP protocols when they become available in the marketplace.

Can We Justify a Transition to GOSIP Now?

Some government agencies say that they will not transition to GOSIP until it provides full functionality. However, if the agency only requires the functionality that is currently provided by GOSIP, then it makes sense to plan to transition

By developing its applications based on GOSIP protocols, it will not be necessary to modify the applications in the future to make use of the GOSIP functionality.

to GOSIP now. If, on the other hand, the agency's applications are not yet supported by existing GOSIP protocols, then it makes little sense to attempt to transition at the present time. Thus, the question is not whether or not to wait until GOSIP provides full functionality (since OSI is continuing to evolve, there always will be additional functionality to be added to GOSIP), but whether the functionality required by the agency is supported by GOSIP now.

An important issue is whether the planned upgrades of GOSIP will be upwardly compatible. No agency will want to throw away its investment in current GOSIP products when new GOSIP versions are introduced. (Unfortunately, this is a real concern, as witnessed in the release of Manufacturing Automation Protocol [MAP] Version 3.0, which was incompatible with the earlier Version 2.1.) It is also important that the cost of upgrading be reasonable. However, if the future versions are upwardly compatible, then the agency will be able to budget the replacement of portions of its system on a scheduled basis.

A final transition question that must also be addressed is the cost of *nut* transitioning to GOSIP now. The cost associated with transitioning is not just the cost of replacing current protocols with GOSIP protocols. It is also the cost associated with modifying the agency's applications to use the additional functionality of the GOSIP protocols. Thus, if an agency is planning to develop a new system, or to carry out a major upgrade of its current applications, then it will be less expensive to transition to GOSIP now, rather than waiting for "full" functionality in the future. By developing the new system using GOSIP protocols now, there will be no need to transition to GOSIP in the future. In addition, by developing its applications based on GOSIP protocols, it will not be necessary to modify the applications in the future to make use of the GOSIP functionality.

Summary and Conclusions

The impact of FIPS 146 on government agencies can be minimized by making sure that they understand the importance of realistic transition planning. Most importantly, each agency's transition plan must be tailored to its own environment. The agency cannot simply copy another agency's transition plan (as a number of agencies did with the Defense Data Network [DDN] transition plan).

The transition plan must provide the agency with the capabilities that it will need to carry out its day-to-day functions. It must be realistic in terms of implementation timeframe and resources required, and it must address the methodology to be used to transition from existing protocols to GOSIP protocols. Thus, for example, it would not make sense to specify the use of application gateways in a tactical environment where there is a limited bandwidth, and where the loss of the gateway will lose the connectivity needed to function on the tactical battlefield.

For agencies with existing networks, the plan should allow for coexistence with existing protocols in the near term, and provide a smooth transition to a full GOSIP implementation in the future.

For agencies with existing networks, transition plans should not have to specify the use of a pure GOSIP stack. The objective of transitioning to GOSIP is to provide for interoperability. This can most readily be accomplished through the use of common application layer protocols, Thus, as a transition strategy, it makes sense to run GOSIP hosts over non-GOSIP networks. For example, by using RFC 1006, GOSIP application layer protocols will be able to run on top of the DOD TCP/IP protocols.

A great fear of any organization is to be placed in the position of being a "guinea pig" for new technology. It does not want to have to learn through trial and error, the correct way to transition to GOSIP. Thus, to minimize the risk inherent in transitioning to GOSIP, it is necessary that MITRE help the government agency understand all of the transitioning issues and ensure that the agency's transition strategy is appropriate for its environment, technically sound. and viable at the present time. For agencies with existing networks, the plan should allow for coexistence with existing protocols in the near term, and provide a smooth transition to a full GOSIP implementation in the future.

About the Author

Richard Nieporent is a lead engineer in the Intelligence and Commander in Chief (CINC) Support Systems Division. He received his A. B. in Physics from Columbia University, and his M. S. and Ph.D. in High Energy Physics from Rutgers University. Richard has been working in the area of communications for the last 20 years. He teaches graduate courses in Data Communications Networks, Local Area Networks, and Computer Network Architectures and Protocols at The Johns Hopkins University. In addition, he is a member of the Computer Science Program Committee, and he is responsible for the Data Communications Concentration for the masters degree program. Richard is also an active member of the ANSI X3T5 Standards Committee that is responsible for the U.S. contributions to the OSI Reference Model and Upper Layer Protocols. At MITRE, Richard has been working on the communications aspects of the Army Tactical Command and Control System. Additionally, he is the chairman of the SOS (GOSIP) cluster group.

Frame Relay: A Next Generation Packet Service

Isadore Schoen

Local area networks (LANs), image transfer applications, and more sophisticated bridges and routers are generating increased amounts of bursty data for transmission over long distances. These systems have outstripped conventional wide area networks (WANs) based upon X.25 and Tl lines. To meet the demand for WANs that can accommodate these systems, fast packet switching techniques have been developed. Fast packet switching is characterized by an abbreviated header, little or no congestion control, and no error control in the network.

One specific implementation of fast packet switching is the Frame Relaying Bearer Service. Frame relay is a high-performance data transfer service that is intended for data communications at rates of up to 2 Mbps using existing physical interfaces. It introduces shorter delays than X.25, provides out-of-band call control, and is part of a true international standard, CCITT 1.122.

Frame relay is one of several techniques labeled "fast packet switching." Today's networks provide reliable end-to-end digital connectivity, with extremely low error rates (bit error rates on the order of 1×10^{-10}). Interest in fast packet switching has developed as new applications have outstripped the capabilities of traditional X.25 networks. Many organizations have implemented dedicated-line technologies, saddling themselves with high costs while sacrificing the flexibility that a packet switched network provides. A new service has been developed that can provide better performance for a reasonable cost. It is expected that frame relay will provide data communication rates comparable to T1 rates (1.544 megabits per second [Mbps] in the U.S., 2 Mbps in Europe).

Frame relay is a packet mode bearer service that provides for the transparent transfer of user information in a packetized manner over a logical link at the user-network interface. Frame relay is an emerging standard described by International Telegraph and Telephone Consultative Committee (CCITT) Recommendation I. 122, "Framework for Providing Additional Packet Mode Bearer Services" and within American National Standards Institute (ANSI) T1.606, "Frame Relay Bearer Service — Architectural Framework and Service Description." It provides connection-oriented packet mode service, link layer multiplexing, and out-of-band call control. It is intended to provide communications for a wide variety of data applications, including:

- Block interactive data applications such as high resolution graphics and images. This type of traffic requires low delays and high throughput.

- Large file transfers, where transit delay is not as critical as in applications requiring block interactive data. High throughput may be required to produce reasonable transfer times for large files.

- Multiplexed low bit-rates, to provide an economical access arrangement for a large group of low bit-rate applications.

- Character interactive traffic such as keyboard-based text editing.

Frame relay is one of several fast packet switching techniques. In frame relay, variable length frames are used to carry data across the network. Another fast packet technology, Asynchronous Transfer Mode (ATM) using cell relay, is also under development. In cell relay, fixed length cells are used to carry data rather than variable length frames over a fiber optic network. Fixed cells can be handled more quickly than variable length frames. Broadband ISDN (BISDN) and Switched Multimegabit Data Service (SMDS) both utilize cell relay techniques. BISDN will utilize Synchronous Optical Network (SONET) access and ATM switching, while SMDS will utilize Institute of Electrical and Electronics Engineers (IEEE) 802.6 switching and DS- 1 or DS-3 access (see figure 1). Both BISDN and SMDS promise higher data rates than frame relay, but product and standards development lags behind frame relay. It is unlikely that BISDN or SMDS will be available for several years, making frame relay more attractive for the present.

Differences Between Frame Relay and X.25

Frame relay differs substantially from X.25. For instance, X.25 provides what may be considered excessive error correction and flow control mechanisms in today's networks. In frame relay, many of these functions are accomplished at the user-to-user interface rather than by intermediate switches. In X.25, the network provides flow control and error correction. In contrast, the frame relay network supports only a minimal set of functions referred to as the core functions. Core functions include frame delimiting, multiplexing, and error detection. These functions simply ensure that frames with the correct address and format are successfully delivered. All other OSI Layer 2 functions, including error recovery and flow control, are the responsibility of users on an end-to-end basis (see table 1). In X.25, multiplexing is accomplished through the use of logical channels at the network layer. Frame relay provides switching at the link layer through the statistical multiplexing of different data link connections on the same physical channel, using the Link Access Protocol on D Channel (LAPD). These differences permit less per-frame processing by the network, resulting in lower delays and higher throughput than X.25 can achieve.

Separation of User and Control Information

Fundamental to ISDN is the principle of separation of user data and control information. The objective of this separation is to integrate fully the control procedures for all services, resulting in one set of protocols for call control across all telecommunications services. This separation allows for the integration of voice and data services by separating the service from the signaling required to request, maintain, and terminate these services. All signaling is carried out-of-band, in the control or C-plane, on a separate D channel. This separation permits the B channel to be used exclusively for user data. Because the B channel is not carrying control data, it becomes simpler—allowing more of the control logic to be implemented in the hardware. Unlike X.25, which carries control data in the same stream as user data, frame relay separates this information into control

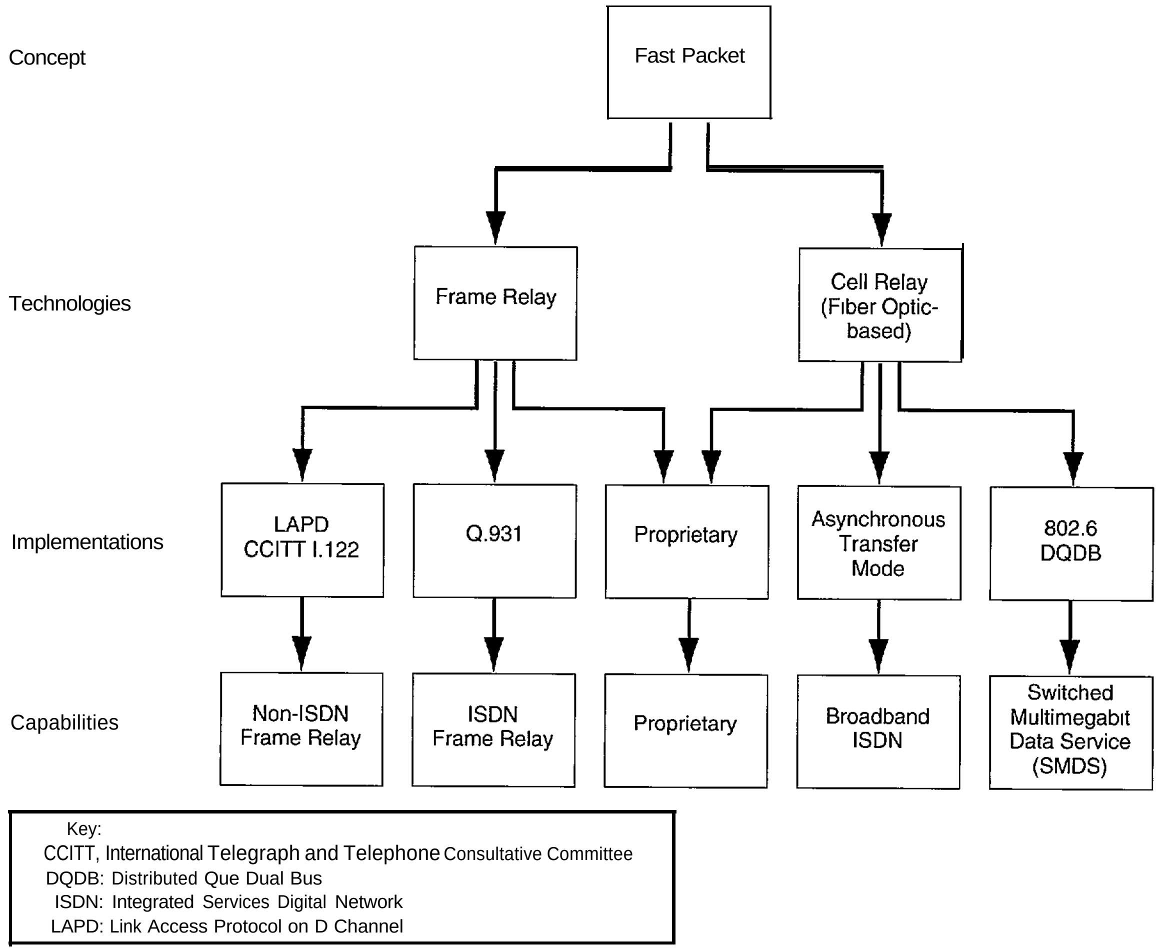


Figure 1. Fast Packet Concept
Frame relay and cell relay represent two technologies under the "fast packet" concept. Each technology provides different capabilities for end users.

(C-plane) and user (U-plane) information. C- and U-plane separation may occur in one of two ways: (1) on another channel (time slot) within the same interface, or (2) on a separate logical link within the same D channel. Out-of-band signaling is commonly utilized in circuit-switched networks.

Using the D channel, frame relay supports both virtual calls and permanent virtual circuits. For virtual calls, dynamic establishment and release of a virtual call are accomplished using a common set of protocols as in X.25 services. Parameters are negotiated at call setup time. The C-plane carries all capabilities for call establishment and termination over the D channel. After call establishment, all transfers of data occur in the U-plane. The U-plane bearer service provides bidirectional transfer of frames, preserving their order during transmission. It detects transmission format and operational errors (e.g., frames

Table 1. Abbreviated Comparison Between Frame Relay and X.25

FUNCTION	X.25	FRAME RELAY
Flag Recognition/Generation	X	X
Transparency	X	X
FCS Checking and Generation	X	X
Recognize Invalid Frames	X	X
Discard Incorrect Frames	X	X
Address Translation	X	X
Fill Inter-Frame Time	X	X
Buffer Packers Awaiting Acknowledgement	X	
Acknowledge Received I-Frames	X	
Generate Reject	X	
Tally Retransmissions	X	
Act Upon a Reject	X	
Respond to Receiver Ready/Receiver Not Ready (RR/RNR)	X	
Detect Out-of-Sequence Packets	X	
Manage Network Layer RR/RNR	X	

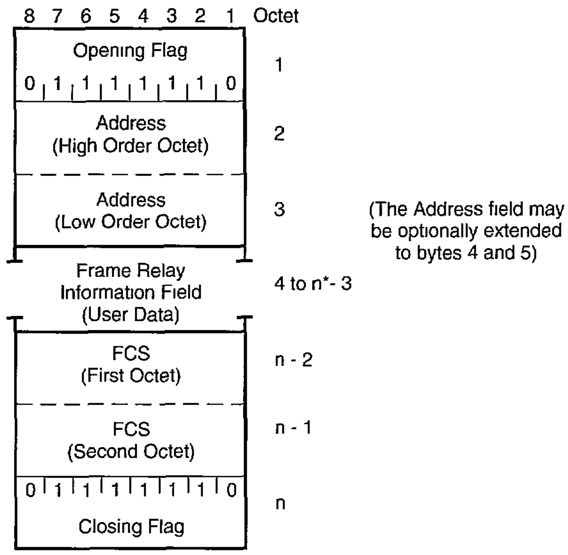
of unknown or incorrect address). It transports the data transparently to the end user. In frame relay permanent virtual circuits, the establishment and release of a connection is accomplished via administrative procedures, with parameters agreed upon at subscription time.

Frame Format

Figure 2 shows the frame relay format as proposed in ANSI Tl .6ca, "Core Aspects of Frame Relay Protocol For Use with Frame Relay Bearer Service." (The bit designated "bit 1" in figure 2 is transmitted first; bit 8 is last.) All frames start and end with a flag sequence consisting of one 0 bit followed by six contiguous 1 bits and another 0 bit (Hex 7E). The flag preceding the address field is called the opening flag; the flag following the frame check sequence (FCS) is referred to as the closing flag. (The closing flag may also serve as the opening flag for the next frame, resulting in "concatenated" frames.) The Address field follows the opening flag, and consists of at least two octets; it may be optionally extended to four octets.

The Frame Relay Information field follows the address field. This is the User Data field, and must consist of an integral number of octets. The default size of this field is 262 octets, with other maximum field sizes negotiable by networks and users. It is expected that many networks will support maximum sizes of up to 4,000 octets. Following the Frame Relay Information field is a 16-bit FCS that validates the integrity of the frame. The FCS is followed by the closing flag.

The Address field is further broken down as shown in figure 3, Address Field Format. This field contains the Data Link Connection Identifier (DLCI). The DLCI is used to identify the logical connection with which a frame is associated. All frames within a particular physical channel that have the same DLCI are associated with the same logical connection. DLCIs have only local significance. The DLCI is bound to routing information at intermediate nodes, thus forming a logical path. These bindings are established at call setup, and are dissolved at call termination. The DLCI is purely a Layer 2 concept, and is not known by Layer 3.



*-n is the maximum field size, either default or as negotiated

Figure 2. Frame Relay Frame Format
A frame is preceded by an opening flag, followed by the address field, information field, FCS, and a closing flag. The closing flag may also be utilized as an opening flag for the next frame.

The Address field also contains several bits referred to as address field variables. These bits are:

- The Address Field Extension (EA) bit, used to indicate whether the current octet is the final octet (when set to 1) in the address field. This allows the address field to vary in length from two to four octets.

- The Command/Response (C/R) bit is reserved for use by applications. It is not used by the frame relay protocol. When a Command frame is to be sent, the C/R bit is set to 0. When a Response frame is to be sent, the C/R bit is set to 1.

- The Backward Explicit Congestion Notification (BECN) may be set by a congested network to notify the user that congestion avoidance procedures should be initiated. This bit is set to 1 to notify the receiver that any frames it transmits may encounter congestion.

- The Forward Explicit Congestion Notification (FECN) is set to 1 to notify the receiver that any frames it receives have encountered congestion.

- The Discard Eligibility (DE) bit which is set to 1 if a frame should be discarded in preference to other frames in congestion situations, i.e., where frames must be discarded to ensure safe operation of the network, and to maintain the expected level of service.

The BECN and FECN are used to adjust transmission parameters at user-network interfaces.

Messages for Frame Relay Connection Control

As previously mentioned, frame relay service utilizes the common signaling protocol used with ISDN (see ANSI Standard T1.607, "Digital Subscriber Signaling System No. 1 [DSS1] — Layer 3 Specification for Circuit Switched Bearer Service"). Frame relay utilizes a subset of the messages defined by T1.607 for connection control. These messages are divided into three categories: call establishment, call clearing, and miscellaneous messages (see table 2). These messages are passed through Layer 2 to the user application, which is responsible for maintaining, forming, and terminating call.

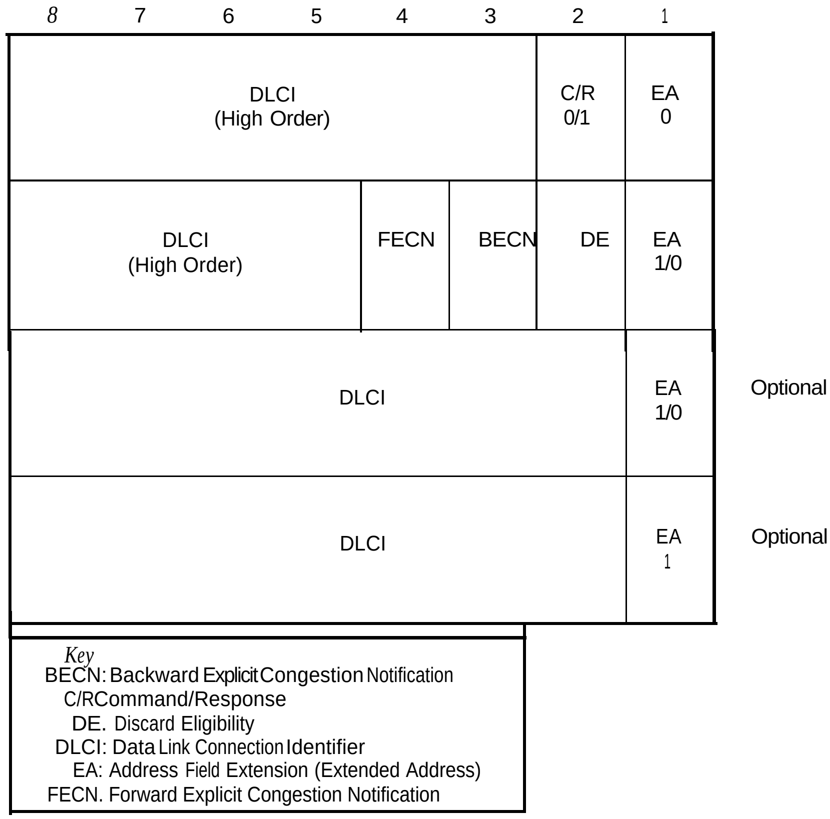


Figure 3. Address Field Format
The address field of a frame may be extended to a total length of 4 octets.

Table 2. Frame Relay Connection Control Messages
Frame relay utilizes a subset of the messages defined by T1.607 for connection control. These messages pass through Layer 2 to the user application.

CALL ESTABLISHMENT	CALL CLEARING	MISCELLANEOUS
Alert Call Proceeding Connect Connect Acknowledge Progress Setup	Disconnect Release Release Complete	Status Status Enquiry

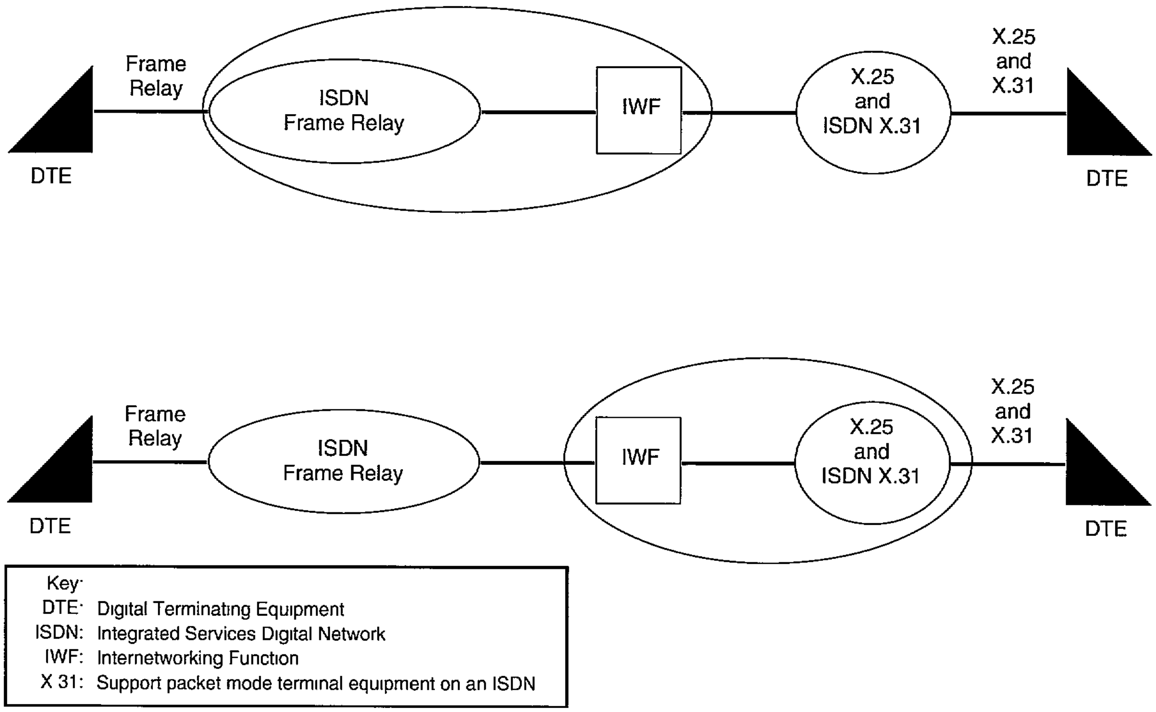


Figure 4. Logical Diagram of IWF
The interworking function can resolve differences between a frame relay network and an X.25 network.

Internetworking

All ISDNs may not provide frame relay so arrangements must be made to provide service between an ISDN offering frame relay service and a packet mode service offered by either an ISDN (X.31) or public switched data network (X.25). Frame relay and X.25 are very different; they offer different services and support different numbering plans. Conversion must be done at either the frame relay network or the X.25/X.31 network.

To provide interconnectivity, CCITT Recommendation I.122 defines an Interworking Function (IWF) that can resolve the differences between a frame relay network and an X.25 network. The IWF would logically be a part of either the frame relay network or the X.25 network (see figure 4). When the IWF belongs to the ISDN, call control mapping takes place. When the IWF belongs to an X.25- or X.31-based network, call control mapping or port access may be used. If the port access method is used, existing X.25 call control procedures are used for the control of virtual circuits.

Conclusion

Frame relay service offers users dedicated-line speeds in a multipoint network while maintaining existing user interfaces. Although future technologies such as BISDN or SMDS promise higher throughputs and lower costs, these technologies will not be available for several years. As more sophisticated applications demand more bandwidth, the need for higher speed services will increase, particularly in the arena of LAN interconnections. No other existing alternative offers the flexibility and cost-effectiveness of frame relay service.

References

1. CCITT Recommendation I. 122, 1988, *Framework for Providing Additional Packet Mode Bearer Services,* Blue Book, ITU, Geneva.

2. ANSI Standard T1.606, *Telecommunications - Frame Relay Bearer Service -Architectural Framework and Service Description.*

3. ANSI Standard T1.6ca, *Telecommunications - DSSI - Core Aspects of Frame Relay Protocol for Use with Frame Relay Bearer Service.*

4. ANSI Standard T1.6fr, *Telecommunications - DSSI - Signaling Specification for Frame Relay Bearer Service.*

5. Johna, J. T., December 1990, "Frame Relay: Changing the Wide-area Services Picture," *Data Communications,* Vol. 19, No. 16, pp. 56-60.

6. Lippis, N., July 1990, "Frame Relay Redraws the Map for Wide-Area Networks," *Data Communications,* Vol. 19, No. 9, pp. 80-94.

7. Lai, W. S., February 1989, "Packet Mode Services: From X.25 to Frame Relaying," *Computer Communications,* Vol. 12, No. 1, pp. 10-16.

8. Lai, W. S., January 1989, "Node and Nodal Architectures for the Interworking Between Frame Relaying Services," *Computer Communication Review,* Vol. 19, No. 1, pp. 72-85.

About the Author

Isadore Schoen received his B.S in Electrical Engineering & Applied Physics from Case Western Reserve University and is currently working towards his M.S in Engineering Management from Old Dominion University. After working for NCR and Diebold, Izzy joined MITRE in 1987. He is now working on the U.S. Courts Data Communications Network within the Information and Communications System Department. He is a member of the IEEE Computer Society, IEEE Communications Society, Amateur Radio Relay League, and Amateur Radio Club.

Concept for a Telecommunications Network Designer's Associate

Kenneth J. Wright

The Designer's Associate is envisioned as an intelligent computer aided design tool for the topological design of telecommunications networks. The system concept features the use of autonomous designers under the direction of a knowledge-based design manager that proposes designs to the user. The designers attack portions of the overall design problem through a cooperative control mechanism and communicate via the blackboard paradigm. Exact and heuristic algorithms, based upon established engineering models, are employed by the designers to modify a prototypical design to meet specific functional and performance requirements for an application. The architecture proposed for the Designer's Associate lends itself to a distributed processing approach and may be implemented with the aid of parallel processing methods.

This paper describes an aspect of software development efforts being conducted within MITRE's Telecommunications Specialty Group. This effort is being directed at the design and implementation of advanced automated design tools for engineers and analysts who design private telecommunications networks and virtual private networks. As part of a more comprehensive program to develop an integrated telecommunications network modeling environment, this particular effort is concentrating on the construction of an "intelligent" software tool the Designer's Associate.

The Designer's Associate will be a knowledge-based system that participates, as an active member, in telecommunications network design teams. Its role in a design team will be to:

- Perform specific functions that are most appropriate for performance by an automated system-for example, maintaining information about the state of the design-in-process or executing computations in support of mathematical modeling.
- Augment the technical skills of other members of a design team by making available and applying knowledge about successful previous designs, about the tools used to create designs, and about the design process itself.

The background section of this paper summarizes the current state of intelligent design research and identifies several recent examples of how knowledge-based systems have been applied to the design of telecommunications networks. The subsequent sections discuss the operational concept for the Designer's Associate and describe the proposed architecture for implementing the system.

Background

Researchers in both industry and academia are exploring new opportunities to apply emerging Computer Aided Design (CAD) tools and techniques to real-world engineering practice. The most familiar and widely publicized examples of this technology are the workstation-based CAD products that help a designer use models and procedures through a sophisticated graphical user interface. Such products offer an easy to use, intuitive means for an engineer to access a set of standardized models for a specific type of problem such as stress analysis of building structures. More advanced versions of these products integrate several types of modeling and analysis tools to address "vertical applications," e.g., airframe design, which in turn employs such methods as structural engineering and composite materials analysis.

Part of the appeal of CAD products stems from the fact that they relieve the designer from tedious or repetitive subtasks such as managing data or doing calculations. In most conventional systems, the human user (engineer) does the designing, not the software tool. The success of the design effort relies upon the judicious use of the automated tools by the engineer, which implies that the engineer possesses some measure of expertise or knowledge about the design process and can recognize what constitutes a "good" design.

Current research seeks ways to allow the computer to perform a greater share of the actual design function, to include the production of candidate designs by fully automated procedures. Some of the motives behind these efforts were stated succinctly by Mostow [7]: "Mechanizing design moving the design process into the machine offers to improve both cost and reliability. To the extent that design can be automated, the productivity of human designers can be enhanced. To the extent that assumptions involved in design can be explicitly represented and automatically enforced, design errors resulting from violated assumptions can be avoided."

One factor influencing the growth of activity in this research area in the past decade has been the modest success achieved in the subfield of artificial intelligence (AI) that deals with knowledge-based systems (KBSs) and, in particular, expert systems. AI researchers have learned some valuable lessons concerning the types of applications that are amenable to the use of existing KBS technology and concepts. Automated design applications, particularly those associated with a recognized engineering discipline, are regarded as potential high-payoff targets that could benefit from the application of methodologies and techniques emerging from applied AI research.

Much of the current AI research in intelligent design is proceeding along two distinct paths. One research direction seeks a comprehensive model of design that can characterize all types of design activity in a unified framework. The other research direction seeks to examine in detail individual aspects of different design activities in various disciplines, with the goal of developing useful insights into how these specific activities can be automated. This latter approach denies the necessity of a grand unified theory of design as a prerequisite for developing useful intelligent design aids.

Although some limited progress has been made in developing general models of the design process, this research is still in its infancy. Remarks published by Mostow [7] in 1985 are still apropos: "The study is still at an early stage in that much of the design process is still poorly understood, let alone automated. We are still developing our models of the design process." The recent paper by Takeda, et al. [11] describes an approach that focuses on integrating cognitive and descriptive models of design processes into a foundation for a computational model. The computational model is used to develop formalisms by which a computer can perform design.

A different approach, being pursued by Chandrasekaran [2] et al., concentrates on a limited class of design activities exhibiting certain properties that can be successfully modeled as a type of search problem. This approach emphasizes the structural aspects of design as a complex of organized tasks. It exploits the task structure of a design activity in order to make the search problem manageable enough to be automated. As explained later in this paper, this class of design activities includes those addressed by the Designer's Associate.

An important classification of design activities that has received some broad acceptance in the AI community gauges the extent to which originality is a factor. As presented in [3], the classification may be defined as:

- Routine design - Performed within a definite state space of possible designs, but resulting designs lie within a relatively small, "normal" region of the state space. Design variables assume values within limited, "normal" ranges.

- Innovative design - Performed within a definite state space of possible designs, with resulting designs occurring outside the "normal" region. Design variables take on values outside "normal" ranges.

- Creative design - Performed in region that is an extension to the known state space of possible designs. Design involves variables not previously encountered.

Current paradigms for knowledge representation and the control of problem-solving processes are much more mature for routine design than for innovative or creative design applications.

Deductive modes of reasoning have been extensively studied by the AI community. Relatively little applied research has been directed at the abductive and circumscriptive types of reasoning [11] that are likely to be involved in non-routine design processes.

The lack of an all-encompassing, unified theory of design has not deterred applied AI researchers from using viable AI models for specific types of design activities. Articles discussing the implementation of knowledge-based systems for design tasks in the telecommunications field have begun to appear in the AI literature. AT&T is using a system named Nestor (discussed in an article by Harris, et al. [6]) to assist engineers in designing packet switched network access configurations in response to user change requests. The system uses real-time access to a corporate network database to obtain information about facilities available at the user site and network utilization. It applies knowledge about the design process to invoke automated procedures to solve design subproblems. A rule-based expert system, Nestor has reportedly improved significantly the productivity and accuracy of the engineering design group it supports.

Researchers at the Dutch PTT developed a knowledge-based system (discussed in van Liempd, Velthuijsen, and Florescu [12] to design data communications user access configurations. The system was designed to assist human network designers in the selection of physical links and equipment. Their approach integrates rule-based and procedural knowledge in a distributed processing environment.

A third example application in telecommunications design is a prototype expert system named ELAND (described in Ceri and Tanca [I]) that addresses problems associated with configuring a local area network (LAN). The system was designed to accomplish three design functions: transform user requirements into system requirements, match system requirements with available commercial products, and assign data files to LAN file servers. Like the other example systems, it uses a rule-based paradigm.

Designer's Associate Concept

The Designer's Associate can be regarded as an integrated set of design tools to assist a network design team in the routine design of private network topologies. The design problem can be summarized as follows. A large, geographically dispersed organization requires various telecommunications services for users at a set of prescribed locations. A minimum cost network of nodes and links must be designed to provide service to these locations subject to a set of known functional and performance constraints. In its most complex form, the problem may involve requirements for a hierarchy of local, metropolitan, and wide area networks to deliver integrated services.

In many real-world telecommunications applications, the design solution space is restricted to design elements that are commercially available and may be limited further to services and equipment that are available under established commercial contracts. In such cases, innovative or creative solutions are neither appropriate nor necessary. The Designer's Associate is intended to support these types of routine design applications.

The system is termed an "associate" because it would participate in the design process along with, and as a partner of, human designers. It would contribute to the design team by performing many of the tedious subtasks involved in the design process and by ensuring that the design team takes into account all available alternatives for design elements.

System Functions

Figure 1 depicts the basic concept of the Designer's Associate. In brief, the user of the system will oversee the network design process, describe functional and performance requirements of the network to be designed, and evaluate designs generated by the system. The system will generate an initial design that fulfills requirements, and then iteratively refine the design in response to the user's feedback. In the course of producing or refining the design, the Designer's Associate will make use of a suite of automated procedures that solve specific design subproblems, e.g., finding constrained minimum spanning trees for a set of network nodes.

In performing its design functions, the system will apply its knowledge base, which incorporates:

- Design Prototypes - A set of predefined design cases representing abstractions of previously successful design exercises, which are used as starting points for generating initial designs.
- Model-Based Algorithms - Formal procedures, derived from telecommunications engineering models, that apply either exact or heuristic methods to optimize design variables.
- Design Process Knowledge - Information about the history and current state of the design task in process, together with guidance for executing the tasks comprising the design process.
- Domain Knowledge - Information about the design elements that are available to construct a network design (such as available product specifications, available services and associated tariffs, etc.), together with rules for how design elements can be assembled.

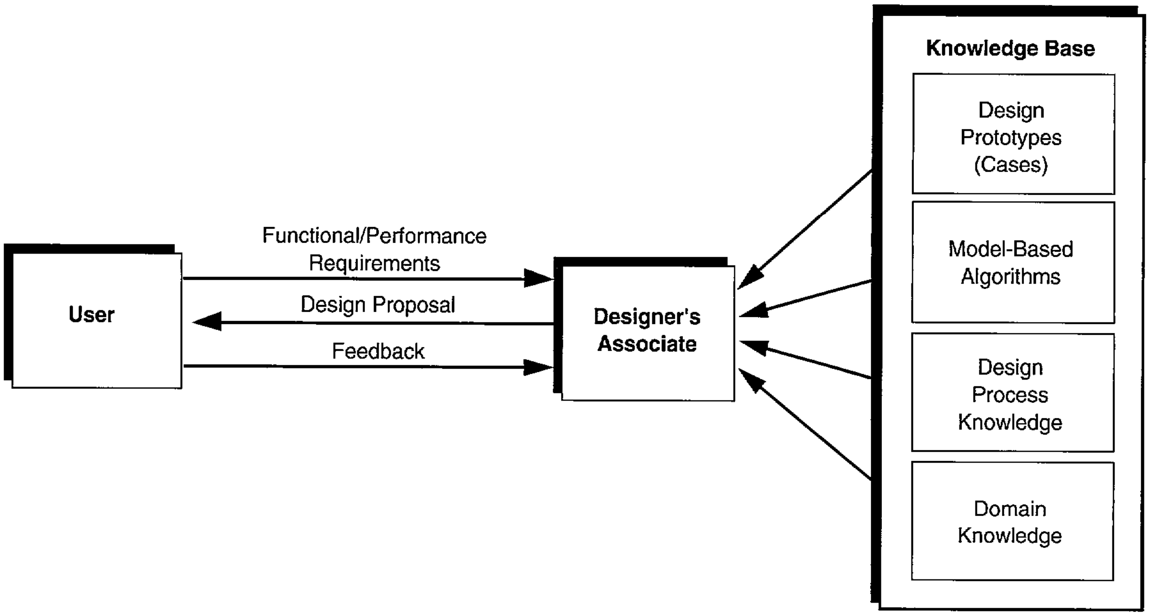


Figure 1. Designer's Associate Overview
The Designer's Associate will interact with the user to establish an initial design and then refine the design through a cyclical process.

The operation of the system will be based upon a cyclical process belonging to the Purpose-Critique-Modify (PCM) family of methods described in [2]. The process is depicted in figure 2. The system executes tasks to accomplish the Propose phase, and produces a design proposal for the user to review. The user critiques that design proposal and identifies to the system any shortcomings that have been found. The system then executes procedures that modify the design in response to the criticism and then enters the Propose phase again. The cycle continues until either: (1) a design proposal is produced that evokes no criticism from the user, or (2) the system exhausts its knowledge about how to modify a pending design proposal and has not yet resolved all shortcomings. As the design progresses through the PCM cycles, it is refined, for example, to account for constraints that were not specified in the original requirements.

Several strategies are available for generating the initial design in the first PCM cycle. The approach being investigated for the Designer's Associate is to apply case-based

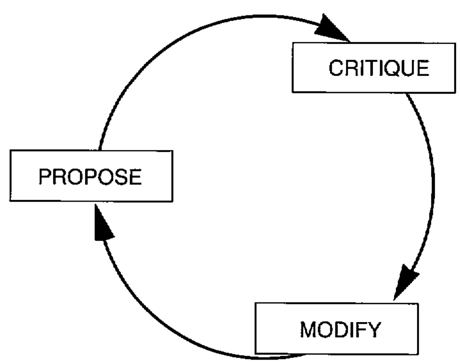


Figure 2. Propose-Critique-Modify Cycle
A design process with three phases. The system proposes a design and the user critiques it. Then the system modifies the design and a new cycle begins.

reasoning methods to match user-specified requirements against a set of prototype network designs. (The design prototype schema for knowledge representation is discussed in [3] and [4].) The prototypes are essentially skeletal descriptions of classes of design elements that can be used to build a complete design fulfilling the requirements specification for the final product. Prototypes used in the Designer's Associate could range in complexity from a complete private network (e.g., a generic packet switched network) to individual network building blocks (e.g., a terminal concentrator configuration). As explained below, prototypes in the system's repertoire will be used on the basis of their suitability for a specific set of design requirements.

In response to the initial requirements described by the user, the system will develop an initial design proposal following the task structure outlined in figure 3. The system will search the available set of design prototypes in a hierarchical manner, attempting to match user requirements as completely as possible with a prototype at the top level before moving down to the next level and attempting to match with prototypes of components from which a design would have to be synthesized. This case-based reasoning approach attempts to develop a design for a given problem by starting with a solution to a similar problem, based on the assumption that the stored solution (the design prototype) will, in some sense, be nearly appropriate for the given problem as well. If at least one prototype match is found, then the closest match will be selected as the basis for the initial design and instantiated; otherwise, the system will request that the user relax some of the problem requirements and the search process will be repeated.

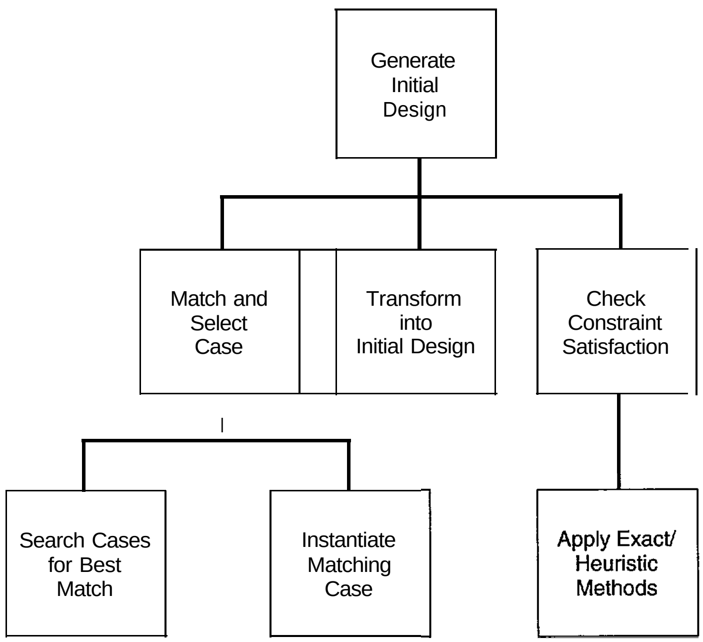


Figure 3. Initial Design Task Structure
To begin the design process, the system will find a design prototype that most closely matches requirements and will transform the prototype into an initial design proposal.

The product of the matching task is a collection of design object instances-a skeletal design structure-that must be "fleshed out" by assigning suitable values to each object's attributes. The system will apply a series of transformations, based on knowledge about the selected design prototype, to achieve the functional requirements of the design problem being solved. For example, a design object instance for a T1 multiplexer configuration might have an attribute named "data-rates-supported" that could be assigned the value "4.8;9.6."

The system will apply the performance requirements of the problem being solved, which may involve constraint satisfaction and, in many cases, optimization of design variables. To accomplish this, the system will execute software that carries out appropriate methods (exact or heuristic) that are based on accepted engineering models. For example, at some stage in the design process the system may be required to find the minimum cost layout of tie trunks to be leased from a local carrier. To solve the problem, the system might set up a mathematical programming problem, execute an appropriate software package, and incorporate the results into the design.

Once the system has produced an initial design that meets as many requirements as possible, it will present the design as a proposal to the user. The user will be shown the structure and description of the initial design, along with an indication of any requirements that could not be satisfied, i.e., performance constraints that were violated or cost factors that could not be optimized.

The user will then carry out the Critique phase of the design cycle by reviewing the proposed design to determine if any functional or performance requirements were overlooked in the initial description provided to the system. If the design is deficient in some respect, it may be that some of the initial constraints were inconsistent or unrealistic. The user may determine that the proposal is generally sound but requires fine tuning within a few detail areas. This phase in the process will result in the user either accepting or rejecting the design proposal.

If the proposal is rejected, the user will indicate to the system what changes must be made to the general or specific requirements, or to the solution procedure itself, to achieve an acceptable design. For example, the user might issue guidance of the form: "If link-loading < 60%, then exclude Tl-carrier-links from the local-link-set." The system could then incorporate this rule into its problem-solving behavior as it modifies the design.

The Modify phase of the design cycle will be performed by the system in an attempt to revise the rejected design proposal in response the user's criticism. If only minor changes have been made by the user to the functional or performance requirements, the system will be able to apply its transformation rules and its constraint satisfaction or optimization procedures to the current design proposal. On the other hand, if the changes are extensive, the system may have to search again through the set of design prototypes to obtain a new starting point for the design process. Once all necessary modifications have been applied to the design, the system will enter the Propose phase of another design cycle and the process will continue.

System Architecture

A key factor influencing the Designer's Associate architecture is the method by which the network topology design problem is approached by telecommunications engineers. It is common practice among network designers to decompose the overall design problem into two subproblems: design of the backbone of links and nodes that provide services throughout some area

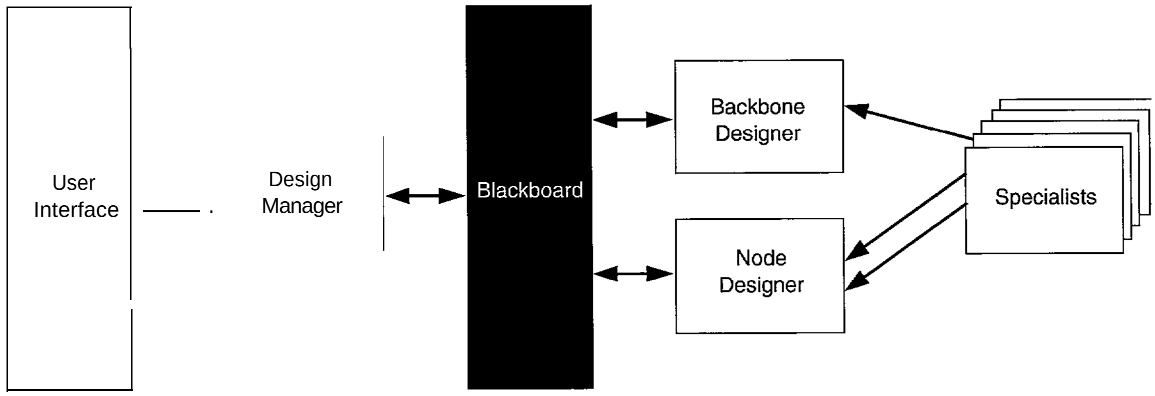


Figure 4. Designer's Associate Architecture
The Designer's Associate communicates with the user through a graphical user interface. A knowledge-based Design Manager proposes designs to the user and directs the activities of a Backbone Designer and a Node Designer. The Designers employ the services of Specialists that solve individual subproblems. Communication among the knowledge systems takes place through the Blackboard.

(i.e., wide, metropolitan, or local area), and design of the node configuration that provides the network user access to the backbone. The hierarchy associated with network design problems is imposed by commercial, legal, and technological factors.

The proposed architecture for the Designer's Associate is depicted in figure 4. The natural structure of the network topology design problem is exploited in this architecture. The bulk of the reasoning and problem-solving (procedural) actions associated with the backbone and node design subproblems will be carried out by two autonomous knowledge systems: one for backbone design (Backbone Designer) and the other for node design (Node Designer). The two Designers will solve their respective design subproblems under a cooperative arrangement to share information about their tasks and results. (Approaches to cooperative problem solving are discussed in [9] and [10].)

The Designers will apply knowledge about solving problems within their domains to:

- Invoke transformation rules to assign values to design object attributes.
- Execute algorithms that employ constraint satisfaction or optimization procedures to assign values to design variables. In the Designer's Associate architecture, these procedures are termed Specialists.
- Exchange results that apply to each other's problem-solving tasks.

The Designers will perform the detail tasks necessary to generate an initial design from a design prototype or to modify a design that has been criticized by the user. They will obtain assistance in these tasks from the set of Specialists that are included in the system. An example of a Specialist would be a Trunk Group Specialist that can be used by the Backbone Designer to size a trunk group through the use of the traffic engineering model appropriate to the type of call disposition method required for the application.

Since the two Designers will operate as peers, a mechanism for coordination and conflict resolution will be needed. This will be the primary responsibility of the Design Manager knowledge system. The Design Manager will:

- Communicate with the system user through a graphical user interface to obtain network design requirements, present design proposals, and obtain criticism on design proposals.

- Perform the search for and selection of design prototypes to establish a starting point for the design process.

- Formulate and issue design tasks to the Designers using knowledge pertinent to the design prototype being used.

- Combine results from the Backbone Designer and the Node Designer into a complete design proposal.

- Reconcile conflicts between the two Designers by relaxing or tightening constraints, assigning priorities to design subtasks, etc.

- Maintain information about the history and current state of the design process.

Communication among the Design Manager and the two Designers will take place through a blackboard structure. (The Blackboard paradigm is described in [8]. An example knowledge system tool that incorporates a blackboard is described in [5].) The Blackboard structure will provide a central data repository for information that must be shared by the Designers and the Design Manager such as control information and solutions to design subproblems. It will also be the storage mechanism for global design information that must be available to all three knowledge systems-for example, design object instances from the design prototype under development.

Under this architecture, design process tasks could be implemented in a distributed fashion with the Design Manager, Backbone Designer, and Node Designer executing on separate processors. In fact, the Blackboard paradigm has been developed expressly for the purpose of building distributed problem-solving systems. Several implementation options can be considered for the Designer's Associate, such as the use of three workstations interconnected by a local area network with the Blackboard implemented on a file server. Another implementation possibility would be to operate the two Designers on a parallel processing system with the Design Manager running on a front-end workstation.

References

1. Ceri, S., and L. Tanca, 1990, "Expert Design of Local Area Networks," *IEEE Expert 5(5),* pp. 23-33.

2. Chandrasekaran, B., 1990, "Design Problem Solving: A Task Analysis," *AZ Magazine* 11(4), pp. 59-71.

3. Gero, J. S., 1990, "Design Prototypes: A Knowledge Representation Schema for Design," *AZ Magazine* 1 l(4), pp. 26-36.

4. Gero, J. S., M. L. Maher, and W. Zhang, 1988, "Chunking structural design knowledge as prototypes," in *Artificial Intelligence in Engineering: Design,* Elsevier Science Publishers B. V., Amsterdam, Netherlands.

5. Gilmore, J. F., S. P. Roth, and S. D. Tynor, 1989, "A Blackboard System for Distributed Problem Solving," in *Blackboard Architectures and Applications*, Academic Press Inc., New York.

6. Harris, J. B., et al., 1989, "Nestor: Network Engineering System for Topology and Requirements," *Expert Systems 6(3).* pp. 134-141.

7. Mostow, J., 1985, "Toward Better Models of the Design Process," *AZ Magazine 6(l),* pp. 44-57.

8. Nii, H. P., 1986, "Blackboard Systems: The Blackboard Model of Problem Solving and the Evolution of Blackboard Architectures-Part I," AI *Magazine 7(2),* pp. 38-53.

9. Smith, R. G. and R. Davis, 1988, "Frameworks for Cooperation in Distributed Problem Solving," in *Readings in Distributed Artificial intelligence,* Morgan Kaufmann Publishers, Inc., San Mateo, CA.

10. Sriram, D., et al., 1989, "Knowledge-Based System Applications in Engineering Design: Research at MIT," *AI Magazine* 10(l), pp. 79-96.

11. Takeda, H., et al., 1990, "Modeling Design Processes," *AI Magazine* 1 l(4), pp. 37-48.

12. van Liempd, G., H. Velthuijsen, and A. Florescu, 1990, "Blondie-111," *IEEE Expert 5(4),* pp. 48-55.

About the Author

Ken Wright joined MITRE's Telecommunications Systems Department in 1989 as a member of the technical staff. He holds an M.S. degree in Operations Research from Case Western Reserve University, where he also received his undergraduate degree. He has provided research and analysis support to federal government agencies since 1975.

Call for Papers

Future Issues of *The Telecommunications Review*

We invite contributions of technical papers for possible inclusion in future issues of *The Telecommunications Review*. The fields of interest within The Telecommunications Specialty Group encompass telecommunications in its many forms. In addition to technical aspects such as transmission, switching, protocols, and related software, papers dealing with network management, costing, introduction of new services, regulatory issues, and strategic planning are welcome. All contributions will be reviewed by the Editorial Board.

To have your paper considered for the next issue, submit five copies by 3 1 October 1991 to:

Henry James, Department Head
Telecommunications Systems Department
The MITRE Corporation
7525 Colshire Drive
McLean, VA 22 102
(703) 883-5404